Asset pricing with complexity

Mads Nielsen[†], Antoine Didisheim

June 5, 2023

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.

JEL classification: G11, G12, G14,

Keywords: Asset pricing; Asymmetric information; Cost of complexity; High-dimensional inference; Return predictability;

I would like to thank my thesis supervisor Norman Schürhoff for his valuable feedback. I also thank Suzanne Vissers and Ahmed Guecioueur for helpful comments.

[†] PhD candidate, Université de Lausanne and Swiss Finance Institute. E-mail: mads.nielsen@unil.ch. Postal: Extranef 225, Quartier UNIL Dorigny, 1015 Lausanne, Switzerland.

1. Introduction

Big data is revolutionizing the finance industry (Goldstein, Spatt, and Ye, 2021) and is the key input to a new information economy (Farboodi and Veldkamp, 2021). Machine learning is recognized as a technology that unleashes the potential of Big Data, but empirical research suggests that there is more to the story (Gu, Kelly, and Xiu, 2020). Applied to return prediction, methods that rely on dimensionality reduction outperform ordinary least squares regression by successfully employing more predictors. However, they are outperformed by neural networks taking advantage of non-linear function approximation. A sophisticated approximation is useful when the true functional form of the underlying prediction problem is unknown. In this paper, I introduce a tractable formulation of such a prediction problem and derive its implications for financial markets.

Investors trade an asset with a pay-off for which some statistical properties are unknown and must be estimated from data. A multitude of data sources is available for designing the statistical model. All data sources would improve investors' predictions of the pay-off if they knew the true model. However, estimating the statistical model generates a loss of predictive performance compared to the baseline, a cost of complexity. Investors limit the cost of complexity by choosing an estimator that optimally trades off bias and variance, given the difficulty of the estimation problem and their level of sophistication, their estimation technology. Estimation technology covers algorithms and data quality/quantity as well as heuristics and experience. An unbiased estimator is generally not optimal due to its high variance. The immediate effect of improving estimation technology may be a larger bias if the trade-off with variance is attractive. Even with the optimal estimator, the cost of complexity of including certain data sources can be so high that it outweighs the benefits, and investors improve their prediction of the pay-off by excluding these data sources.

In a representative agent model, econometricians analyzing the generated market data will find predictability in returns even if they perform their analysis out of sample unless their optimal bias coincides with investors' bias. Furthermore, a lower cost of complexity due to better estimation technology can lead econometricians to consider data sources that investors originally ignored. While this is actual predictability not exploited by investors, it is not due to risk premia or irrational mispricing but the complexity of extending the statistical model to include those data sources.[1] In both cases, the predictability generated by the necessity of function approximation is fundamentally different from the in-sample

[1]For a real-world example, consider the application of natural language processing to newspaper articles and annual reports. Both sources of information have been available and presumably used by investors. However, the ability to search for patterns across thousands and thousands of publications is unique to the statistical/algorithmic approach to textual analysis.

predictability in returns related to parameter uncertainty (Lewellen and Shanken, 2002; Martin and Nagel, 2021). With parameter uncertainty, in-sample predictions of returns are biased, but conditioning only on information available at the time of trading (out-of-sample testing) removes the bias. With function approximation, advances in estimation technology, whether those are new tools like machine learning or better heuristics, become a source of predictability that persists out-of-sample as well.

Embedding the prediction problem in the workhorse asymmetric information model of Grossman and Stiglitz (1980), with informed and uninformed investors, the cost of complexity may be so high that no one would decide to become an informed investor even if data was available for free. The condition for informed predictions to outperform uninformed predictions is related to the condition for whether or not to ignore new data sources, but they are not identical. It is possible for informed predictions to deteriorate due to higher dimensional[2] data without informed investing being given up. I explicitly show how information asymmetry is another source of predictability that does not disappear out-of-sample. Earlier works by O'Hara (2003) and Biais, Bossaerts, and Spatt (2010) analyze asset pricing implications of models similar to the baseline model without the estimation problem, but rather than discussing return predictability per se, they focus on implications for CAPM. In the full model, the interaction between asymmetric information and function approximation produces out-of-sample 'echoes' of in-sample results even if econometricians manage to match the optimal bias and active information set of investors. This problem is particularly pronounced in markets prone to large supply shocks, high levels of noise trading. This is a concern for empirical work since trading noise might appear to be behavioral bias as these markets are populated by investors who can be considered particularly susceptible to such biases, i.e., specific stocks with large exposure to retail investors. The interaction, however, also represents an opportunity for empirical analysis since cross-sectional variation in noise trading affects the predictability arising from investors' optimal bias but not ignored data sources. The channel is price responsiveness, how strongly prices react to information and supply shocks. Generally, shocks that enter the price through investors' predictions will vary with price responsiveness, and others, e.g., priced risk, do not. In either case, additional variation is required to distinguish between such sources of predictability and those generated by the prediction problem.

The model provides additional predictions for price informativeness, price pressure and reversals, trading volume, and fund performance. Objective price informativeness is subject to a bias-variance trade-off, which is not optimized by the solution to the investors' prediction problem. Investors might ignore new data sources that improve price informa-

[2]Such a deterioration can occur if subsets of the data cannot be ignored selectively.

tiveness. Improving estimation technology closes the gap between the private and social value data. Price pressure is generated both by supply shocks and estimation noise of the prediction problem, making short-term price reversals more likely, other things being equal. In contrast to an established tradition of analyzing price pressure (Campbell, Grossman, and Wang, 1993; Hendershott and Menkveld, 2014), conditioning on price variance is more effective than trading volume for distinguishing between the two sources. Fund performance is subject to an unanticipated transfer between investors when comparing investors' ex-ante expectations of profits to the ex-post average of realized profits. The sign of this transfer depends on whether investors over- or underestimate the covariance between the pay-off and informed investors' prediction and appears predictable in retrospect.

In an empirical application of the model, I analyze two patterns of predictability from the literature on predicting the equity risk premium following (Welch and Goyal, 2008). The first pattern emerges *across* studies. A group of predictive variables outperforms the historical mean in the earlier part of the sample, followed by under-performance in the later part. The turn-around falls in the early 1990s. It follows the rise of the private computer in the 1980s and coincides with the early years of the internet. The second pattern emerges *between* studies and is that later papers present estimation approaches that outperform earlier papers (Campbell and Thompson, 2008; Rapach, Strauss, and Zhou, 2010; Neely, Rapach, Tu, and Zhou, 2014; Buncic and Tischhauser, 2017; Hammerschmid and Lohre, 2018). The second pattern does to a certain extent coincide with the introduction of additional data sources. However, I show that the pattern can be replicated by comparing ordinary least squares to regularized linear models (Ridge regression and LASSO). An improvement in investors' estimation technology over the run of the data series can explain the first pattern, and improved estimation technology employed in later studies can explain the second pattern. In the model, technology has a one-dimensional and, as such, ordered representation.

Calibrating the representative agent model to the data to replicate the first pattern, the change in predictability requires a large shift in technology, with the relevant parameter tripled from the earlier to the later period. The calibration is consistent with a substantial improvement in investors' estimation technology. Furthermore, the calibration demonstrates the importance of modeling optimal bias, i.e., a bias traded off for lower variance. The fit that generates the pattern of predictive out-performance followed by under-performance does not produce a decrease in the bias of investors' estimator but rather an increase.

1.1. Related literature

The central prediction problem in this paper is motivated by results in the empirical literature on applying machine learning to asset pricing (Gu et al., 2020; Ma, 2021). The formalization abstracts the practical approach of estimating an information structure with factor loadings and factors as distinct sub-problems formulated for linear sub-problems in Kelly, Pruitt, and Su (2017) and extended to non-linear sub-problems in Gagliardini and Ma (2019) and Gu, Kelly, and Xiu (2021) (in particular Figure 2 of that paper is a clear representation). An emerging literature on the virtue of complex models (Kelly, Malamud, and Zhou, 2022; Didisheim, Ke, Kelly, and Malamud, 2023) establish theoretically and empirically how over-parametrized models achieve good performance in return prediction. Through a rigorous application of random matrix theory the existence of a complexity wedge between a feasible and an ideal model is established in a partial equilibrium setting. The virtue of complex models suggests that increasingly complex models can outperform their antecedents and provides a motivation for studying how technological improvement of prediction methods affects a general equilibrium setting as done in the this paper.

In the analysis of return predictability as econometricians' prediction problem, this paper builds on the literature on learning in financial markets, specifically learning about parameters (Lewellen and Shanken, 2002; Pastor and Veronesi, 2009) and its extension to the high-dimensional regime of big data (Martin and Nagel, 2021). The step from learning to machine learning introduced in this paper is achieved by necessitating function approximation and, as such, introducing a bias-variance trade-off in investors' prediction problem.

The model extends classic models of information aggregation (Grossman and Stiglitz, 1980; Hellwig, 1980; Kyle, 1985) where one signal is sufficient to model. That signal might be the outcome of a complicated process of following news, analyzing company and industry fundamentals, or having private information about a firm, but the sources themselves are not important. The cost of complexity depends on the information structure, which breaks the irrelevance of the individual sources. The setting is also distinct from the multi-asset setup of Admati (1985) in which the relevance of the information structure comes from the multitude of assets rather than information sources.

In the literature on costly information acquisition (Van Nieuwerburgh and Veldkamp, 2010), a feedback effect between trading and learning decisions makes the covariance structure of pay-offs and signals relevant. The cost functions in these specifications are rather flexible and have been modeled on the spectrum from rational inattention, see Sims (2003), to the entire process of cleaning, evaluating, and processing data, see Dugast and Foucault (2020). The cost of complexity in my model is conceptually different from these exogenous cost structures in that it is an integrated part of the estimation problem and is inherently a cost measured in predictive performance. In terms of implications, the function approximation prediction problem is distinct from costly information acquisition in that it introduces an optimal bias. Furthermore, the decision to include a data source or not is not strategic, as in anticipating the feedback between trading and learning, but purely based on the statistical properties of the information. Allowing this decision to be strategic in a model about constrained but optimal prediction would correspond to allowing investors to distort their beliefs against their better knowledge, more like the optimal self-deception of Brunnermeier and Parker (2005) than rational inattention. Ultimately, the burden put on investors' ability to comprehend the full information counterfactual is smaller in my model as the trading and prediction problems are separated. In this way, the model is also distinct from the bounded rationality models of complexity of Gabaix (2014), and Molavi, Tahbaz-Salehi, and Vedolin (2021).

A pertinent question following the rise of big data has been whether it has made prices more informative. Early results in Bai, Philippon, and Savov (2016) suggest that this is the case, whereas later findings suggest that a subset of firms drives earlier results (Farboodi, Matray, Veldkamp, and Venkateswaran, 2020) and that price informativeness of other subgroups have been constant or might even have declined. This discussion has taken place in the equities market space, and the explanation proposed by Farboodi et al. (2020) focuses on the value of data about large firms versus small ones. In the presence of a cost of complexity, more data can lead to worse predictions if investors cannot separate the data into distinct sources and ignore some of them. If investors can separate the data, they might still ignore new data, and, as a result, price informativeness is unchanged. Advances in estimation technology asymptotically close the gap, but the pace of convergence at lower levels depends on the information structure. Therefore, it varies across assets even if the estimation technology is standardized.

In parallel to the discussion of price informativeness, the discussion of return predictability has, in the last ten years, seen the declaration of the 'factor zoo' (Cochrane, 2011), a replication crisis (Harvey, 2017; Hou, Xue, and Zhang, 2020), and a potential rebuttal by reference to a Bayesian baseline (Jensen, Kelly, and Pedersen, 2021). It might not only be econometricians who are affected by the high-dimensional inference problems but also investors (Martin and Nagel, 2021), and investors' solution to their prediction problem feeds into econometricians' empirical analysis. Changes in estimation technology can generate return predictability even out-of-sample, and additional sources of variation are necessary to distinguish it from risk premia and/or anomalies.

I proceed as follows. In Section 2, I show the inference problem in the context of a representative agent model and how it generates optimal bias and a cost of complexity.

In Section 3, I embed the inference problem in an asymmetric information model with heterogeneity across agents. In Section 4, I discuss model predictions and the value of data analytically and numerically. Section 5 covers the empirical aplication to patterns in predictability of the equity risk premium and Section 6 implications for further empirical work before I conclude with Section 7.

2. Representative agent

The inference problem of the representative agent (or of the informed investors in my extension to a Grossman and Stiglitz (1980) setting in Section 3) is my key information friction. It presents a principled deviation from the dogma of rational expectations by only allowing partial knowledge of the true underlying information structure. The remaining part must be estimated, and I assume that this is done by choosing an estimator that optimizes the quality of the prediction of a risky pay-off, with mean squared error as the measure of quality.

To highlight the impact of the main information friction, assume investors are symmetric and have demand that is linear in the difference between their prediction and the price.[3] Trading one risky asset and one risk-less asset in elastic supply in one period and consuming their wealth in the second (see Appendix B.1), market clearing amongst symmetric agents require price to equal prediction. Denoting the pay-off by y and investors prediction of $\hat{y}$ that is $p = \hat{y}$. In Section 3, I embed the inference problem described below in the asymmetric information work-horse model of Grossman and Stiglitz (1980), introducing heterogeneity in agents.

2.1. Inference problem

For clarity, I impose a factor structure on the pay-off $y = \boldsymbol{\beta}^\top \boldsymbol{q}$ and allow for high dimensionality through factors $\boldsymbol{q}$ with signals $\boldsymbol{s}$ that are well-behaved. Meanwhile true factor loadings $\boldsymbol{\beta}$ are constant (and finite) but must be estimated from noisy data by estimator $\hat{\boldsymbol{\beta}}$, which, due to the noise, is a random variable. The choice of estimator is the key problem in the model.

Assumption 1 (Estimator choice)**.** *Investors minimize the mean squared error of their predictor $\hat{y}$ by trading off the bias and variance of the estimator $\hat{\boldsymbol{\beta}}$. The elements of the vectors of biases is assumed to be finite and the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ full-rank. True $\boldsymbol{\beta}$ is constant and element-wise finite.*

[3]Risk neutrality or the demand function described in Appendix B.3 or Appendix B.4 are all valid choices.

Intuitively, Assumption 1 says that investors care about making the best possible prediction, in contrast to putting more weight on unbiasedness for instance. Furthermore, while turning estimator choice into an optimization problem is inspired by the approach to predictions of machine learning the trade-off could be considered a prior selection mechanism. For tractability, an additional assumption on the independence of the noise in the estimator is important.

Assumption 2 (Independent noise). *The noise in the data of estimator $\hat{\boldsymbol{\beta}}$ is independent of factors $\boldsymbol{q}$ and signals $\boldsymbol{s}$.*

The noise in the estimation data makes $\hat{\boldsymbol{\beta}}$ a random variable with unconditional expectation $E\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\mu}_\beta$ and variance-covariance matrix $Var\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\sigma}_\beta^\top \boldsymbol{R}_\beta \boldsymbol{\sigma}_\beta$ where $\boldsymbol{R}_\beta$ is the correlation matrix of $\hat{\boldsymbol{\beta}}$ and I drop the hat on subscripts to avoid clutter.

Factors and signals are well-behaved, assuming joint normality of non-constant, non-redundant variables. I use the notation $\boldsymbol{\Gamma} := \boldsymbol{R}_{qs}\boldsymbol{R}_s^{-1}\boldsymbol{R}_{qs}^\top$ where matrices $\boldsymbol{R}_{qs}, \boldsymbol{R}_s, \boldsymbol{R}_{sq}$ are correlation matrices. Additionally, I denote a diagonal matrix by $\boldsymbol{D}$.

Assumption 3 (Gaussian factor expectations). *Factors and signals follows a multi-variate normal distribution, such that the conditional expectation of factors given signals is*

$$\boldsymbol{\zeta} := E[\boldsymbol{q}|\boldsymbol{s}] = \boldsymbol{\mu}_q + \boldsymbol{\Sigma}_{qs}\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{s} - \boldsymbol{\mu}_s), \text{ where } \boldsymbol{\zeta} \sim \mathcal{N}\left(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_\zeta\right),$$

with element-wise finite expectations $\boldsymbol{\mu}_q$ and full-rank variance-covariance matrix

$$\boldsymbol{\Sigma}_\zeta = \boldsymbol{\Sigma}_{qs}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{qs}^\top = \boldsymbol{D}_{\sigma_q}\boldsymbol{R}_{qs}\boldsymbol{D}_{\sigma_s}\boldsymbol{D}_{\sigma_s}^{-1}\boldsymbol{R}_s^{-1}\boldsymbol{D}_{\sigma_s}^{-1}\boldsymbol{D}_{\sigma_s}\boldsymbol{R}_{qs}\boldsymbol{D}_{\sigma_q} = \boldsymbol{D}_{\sigma_q}\boldsymbol{\Gamma}\boldsymbol{D}_{\sigma_q}.$$

To derive the bias-variance trade-off of Assumption 1, I first decompose the mean squared error of predictor $\hat{y}$. Some notation is helpful here. The second moment matrix of conditional expectation $\boldsymbol{\zeta}$ and unconditional bias of estimator $\hat{\boldsymbol{\beta}}$ are respectively

$$\boldsymbol{\Omega}_\zeta := E\left[\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\right] = \boldsymbol{\mu}_q\boldsymbol{\mu}_q^\top + \boldsymbol{\Sigma}_\zeta, \quad \text{and} \quad \boldsymbol{\varepsilon}_\beta := E\left[\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta} - \boldsymbol{\mu}_\beta.$$

Lemma 1 (Mean squared error decomposition). *The mean squared error of predictor $\hat{y}$ can be decomposed into three terms*

$$\begin{aligned} E\left[(y - \hat{y})^2\right] &= (E[y] - E[\hat{y}])^2 + Var[y] + Var\left[\hat{y}\right] - 2Cov\left[y, \hat{y}\right] \\ &= \underbrace{\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Omega}_\zeta \boldsymbol{\varepsilon}_\beta}_{\textit{bias-squared term}} + \underbrace{\boldsymbol{\sigma}_\beta^\top (\boldsymbol{R}_\beta \odot \boldsymbol{\Omega}_\zeta)\boldsymbol{\sigma}_\beta}_{\textit{variance term}} + \underbrace{Var[y|\boldsymbol{s}, \boldsymbol{\beta}]}_{\textit{irreducible error}} \end{aligned}$$

Proof. Given Assumption 2, the variance of the predictor is the variance of two independent random vectors, which can be written using the Hadamard product $\odot$ (see Appendix A.1)

$$Var[\hat{y}] = \boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta - \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\beta} + 2\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta (\boldsymbol{\beta} - \boldsymbol{\varepsilon}_\beta) + \boldsymbol{\sigma}_\beta^\top (\boldsymbol{R}_\beta \odot \boldsymbol{\Omega}_\zeta) \boldsymbol{\sigma}_\beta, \tag{1}$$

and $\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta (\boldsymbol{\beta} - \boldsymbol{\varepsilon}_\beta) = Cov\,[y, \hat{y}]$ which cancels out with the negative covariance terms of the mean squared error. The squared bias of predictor $\hat{y}$ can be written as a quadratic form of the bias of the estimator and factor means $\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q \boldsymbol{\mu}_q^\top \boldsymbol{\varepsilon}_\beta$ and collected in $\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Omega}_\zeta \boldsymbol{\varepsilon}_\beta$ with $\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta$ of $Var[\hat{y}]$ in equation (1). The remaining two terms can be collected in the conditional variance under the true model

$$\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_q \boldsymbol{\beta} - \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\beta} = Var[y|\boldsymbol{\beta}] - Var[\hat{y}|\boldsymbol{\beta}] = Var[y|\boldsymbol{s}, \boldsymbol{\beta}]$$

Existence of the moments follows from Assumption 1 and Assumption 3. □

The labeling of the three terms in Lemma 1 matches the bias-variance decomposition as it is usually defined (see Hastie, Tibshirani, and Friedman (2009) chapter 7.3), and the variance under the true model is the irreducible noise from the perspective of choosing the estimator $\hat{\boldsymbol{\beta}}$. Therefore, the the bias-variance trade-off for a given information set only applies to the first two terms. This is in contrast to considerations on expanding the information set as I will return to in Section 2.4. Notice that the decomposition of Lemma 1 does not depend on the asumptions on the specific functional form of the bias and variance of the estimator except for the finiteness and full-rank condition of Assumption 1.

2.2. Bias-variance trade-off

More structure is necessary to formulate a meaningful minimization of the first two terms in Lemma 1. One piece is the constraint that it is a trade-off between bias and variance. Imposing the constraint directly on the bias and variance terms neglects the factor structure of the problem. Instead, I specify element-wise symmetric functions for bias $\boldsymbol{\varepsilon}_\beta$ and volatility $\boldsymbol{\sigma}_\beta$ and extend Assumption 2 to assume no correlation *between* factor loadings as well.[4] Bias and volatility functions are linked through a vector of controls $\boldsymbol{c}$. The controls are an abstraction that captures the choices involved in choosing an estimator. They incorporate both the high-level decision of which estimator (e.g. ordinary least square, LASSO, neural net etc.) but also the details such as how to clean the data and tune any hyper-parameters

[4]Estimating factors and factor loadings are two sub-problems of the prediction problem. The assumptions made about the factors and signals of the risky pay-off in Assumption 3 are separate from the specification of the factor loadings.

the estimator may have.

Assumption 4. *Bias and volatility are element-wise symmetric functions* $\varepsilon_{\beta i} = f_\varepsilon(c_i)$ *and* $\sigma_{\beta i} = f_\sigma(c_i)$*, such that* $\partial\varepsilon_{\beta j}/\partial c_i = \partial\sigma_{\beta j}/\partial c_i = 0\ \forall j \neq i$ *and factor loadings are uncorrelated, i.e.* $\boldsymbol{R}_\beta = \boldsymbol{I}$*.*

In this way, the trade-off constraint can be imposed cleanly as a set of pairwise restrictions $f'_\varepsilon(c_i)f'_\sigma(c_i) < 0$, and the interactions between factors loadings arise from the minimization rather than being imposed on it. The structure of these interactions is entirely determined by the factor structured encoded in $\boldsymbol{\Omega}_\zeta$ as the weighing matrix of the variance term simplifies to $\boldsymbol{I} \odot \boldsymbol{\Omega}_\zeta = \boldsymbol{D}_{\Omega_\zeta}$. To motivate Assumption 4, recall that the randomness in the estimator of the factor loadings follows from the noise in the data and (potentially) the estimation method. Without a specific type of interaction between the two in mind, it seems prudent to limit the impact on the solution of structure imposed on the noise of the problem. However, maintaining the assumption of a pairwise trade-off, interactions can be introduced by specifying a correlation matrix $\boldsymbol{R}_\beta$ and simply replacing the diagonal matrix $\boldsymbol{D}_\zeta$ by $\boldsymbol{R}_\beta \odot \boldsymbol{\Omega}_\zeta$ in the following derivations.

Applying Assumption 4 to the first two terms of Lemma 1 and requiring that volatility is non-negative, the bias-variance trade-off as a constrained minimization is

$$\min_{\boldsymbol{c}} \Theta := \boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Omega}_\zeta \boldsymbol{\varepsilon}_\beta + \boldsymbol{\sigma}_\beta^\top \boldsymbol{D}_{\Omega_\zeta} \boldsymbol{\sigma}_\beta \quad \text{subject to} \quad f'_\varepsilon(c_i)f'_\sigma(c_i) < 0, f_\sigma(c_i) \geq 0\ \forall c_i \in \boldsymbol{c}. \tag{2}$$

The strict inequality of the constraint on the product of first derivatives requires that both are non-zero. To derive a closed form solution, I impose additional restrictions on the functional form of bias and volatility.

Assumption 5. *Bias is a linear and volatility an affine function of control given by*

$$f_\varepsilon(c_i) = k_\varepsilon c_i, \quad f_\sigma(c_i) = k_\sigma c_i + k_{\sigma 0}, \quad \textit{subject to} \quad k_\varepsilon k_\sigma > -\infty, k_{\sigma 0} \in (0, \infty).$$

Assumption 5 is more technical than Assumption 4, and its motivation is to limit the problem defined in (2) to a class of minimizations with unique minima, and make it easier to parse equilibrium outcomes. The assumption of affinity of the volatility function is not necessary for uniqueness and has the drawback that while a solution to an unrestricted solution to optimization (2) exists (see Proposition 1) it might not be feasible as it could violate the non-negativity constraint. However, when that solution is feasible, it is available in closed form and feasibility only depends on the factor structure, not the parameters of the bias and volatility functions. Bias can be extended to an affine form without fundamentally altering the form of optimal controls and minimized bias-variance (see Appendix A.2.3) but it comes at the cost of more involved expressions. In contrast, restricting both functions to a linear form leaves controls of all zeros as the only solution to the minimization. Restricting bias rather than variance makes it possible to study the impact of forcing the estimator to be unbiased by setting controls to zero $\boldsymbol{c} = 0$ for an estimator with variance $\boldsymbol{\sigma}_\beta = k_{\sigma 0}\mathbf{1}$.

2.3. Optimal bias

I denote the optimal controls by a star $\boldsymbol{c}^* = \arg\max_{\boldsymbol{c}} \Theta$ and denote the minimized objective as $\chi := \Theta|_{\boldsymbol{c}=\boldsymbol{c}^*}$. In Section 2.4, I show formally that χ is the cost of complexity and it increases in the number of signals n_s. Additionally, to describe the solution to optimization (2) it is convenient to define the ratio of the slope parameters of the bias and volatility functions as $k_c := k_\sigma / k_\varepsilon$. I will refer to the square of the ratio of slope parameters k_c^2 as the estimation technology parameter because the cost of complexity χ is everywhere decreasing in it (see Equation 3). In contrast, I will interpret the constant of the volatility function $k_{\sigma 0}$ as the difficulty of the estimation because it determines the cost of complexity under the inefficient but unbiased estimator $\Theta|_{\boldsymbol{c}=0}$ that is only optimal for $k_c^2 = 0$. Increasing k_c^2 changes the trade-off whereas increasing $k_{\sigma 0}$ simply scales up bias, volatility, and the cost of complexity.

Proposition 1 (Bias-variance trade-off solution)**.** *Under Assumption 4 and Assumption 5, the unconstrained solution to optimization (2) and cost of complexity are*

$$\boldsymbol{c}^* = -k_\sigma^{-1} k_{\sigma 0} \left\{ \boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1} \boldsymbol{X}^{-1} \right\} \mathbf{1}, \quad \chi = \Theta|_{\boldsymbol{c}=\boldsymbol{c}^*} = k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{X}^{-1} \mathbf{1}, \ \textit{where } \boldsymbol{X} = k_c^2 \boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1}$$

This solution to the unconstrained minimization problem always exists and it is unique.

Proof. For the complete algebraic manipulation see Appendix A.2.1. Existence follows from the positive definiteness of $\boldsymbol{X}$, and uniqueness from the positive definiteness of the Hessian matrix of the objective $k_\varepsilon^2 \boldsymbol{\Omega}_\zeta + k_\sigma^2 \boldsymbol{D}_{\Omega_\zeta}$. Since both are sums of positive definite matrices they are also positive definite. □

Corollary 1.1 (Optimal bias and volatility)**.** *Bias and volatility only depend on slope parameters k_ε and k_σ through their ratio k_c*

$$\boldsymbol{\varepsilon}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} = -k_c^{-1} k_{\sigma 0} \left\{ \boldsymbol{I} - \boldsymbol{D}_\Omega^{-1} \boldsymbol{X}^{-1} \right\} \mathbf{1} \geq \mathbf{0}, \quad \boldsymbol{\sigma}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} = k_{\sigma 0} \boldsymbol{D}_\Omega^{-1} \boldsymbol{X}^{-1} \mathbf{1}.$$

Proof. For the proof of the inequality see Appendix A.2.2. □

Proposition 1 leaves open the question of the feasibility of the solution described. In Assumption 6, I present the condition for feasibility of the solution in Proposition 1. Technological developments captured by changes in the technology parameters k_c and $k_{\sigma 0}$ do not affect the status of the feasibility condition because it does not depend on them.

Assumption 6 (Bias-variance trade-off feasibility)**.** *I assume that the following element-wise vector inequality holds*

$$\boldsymbol{\sigma}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} > \mathbf{0} \iff \boldsymbol{\Omega}_\zeta \mathbf{1} > \mathbf{0}.$$

See Appendix A.2.2 for the algebraic manipulation of $\boldsymbol{\sigma}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*}$ that demonstrates equivalence of the two inequalities. Informally, since the on-diagonal entries of $\boldsymbol{\Omega}_\zeta$ are positive, Assumption 6 restricts the set of factor structures with a feasible solution to structures with "not too" negative average cross-second moments of conditional expectations of factors given signals.

For the interpretation of the square of the ratio of slope parameters k_c^2 as a measure of estimation technological development, notice that the cost of complexity is indeed decreasing in it

$$\frac{\partial \chi}{\partial k_c^2} = -k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{X}^{-1} \boldsymbol{\Omega}_\zeta^{-1} \boldsymbol{X}^{-1} \mathbf{1} \leq 0, \tag{3}$$

but in a non-linear fashion, which means that it changes the trade-off between bias and variance rather than simply scaling them up or down. This improvement could represent new techniques or better input data for the estimation of factor loadings $\hat{\boldsymbol{\beta}}$ since both are abstracted into the properties of the estimator in Assumption 1 and Assumption 2.

2.4. Cost of complexity

To demonstrate that the minimized objective χ is the cost of complexity and the variance under true model $Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I]$ is the balancing cost of simplicity, I first provide a recursive formulation of the second moment matrix of conditional expectations $\boldsymbol{\Omega}_\zeta$ based on block matrix inversion and multiplication. The recursion is over the number of signals n_s, and concerns the decision of including the n_sth signal in the vector of signals $\boldsymbol{s}$. The main step is to operate on the matrix of correlations $\boldsymbol{\Gamma}_{n_s} = \boldsymbol{R}_{qs,n_s} \boldsymbol{R}_{s,n_s}^{-1} \boldsymbol{R}_{qs,n_s}^\top$, and in particular the inverse signal correlation matrix $\boldsymbol{R}_{s,n_s}^{-1}$. Explicit derivations are in the Appendix A.3, but it is necessary to introduce some notation here. Denote the correlation between the new signal n_s and the extant signals $\boldsymbol{s}_{n_s-1}$ by $\boldsymbol{\rho}_{s,n_s} := Corr[\boldsymbol{s}_{n_s-1}, s_{n_s}]$, the correlation between the new signal and factors $\boldsymbol{q}$ by $\boldsymbol{\rho}_{q_i s,n_s}^\top = (\boldsymbol{\rho}_{q_i s,n_s-1} \;\; \rho_{q_i s_{n_s}})$, and define the correlation correction $\rho_{s,n_s|n_s-1} := 1 - \boldsymbol{\rho}_{s,n_s}^\top \boldsymbol{R}_{s,n_s}^{-1} \boldsymbol{\rho}_{s,n_s}$. With this notation, it is possible to define the vector $\boldsymbol{\phi}_{n_s}$ with elements $\phi_{i,n_s} := \boldsymbol{\rho}_{q_i s,n-1}^\top \boldsymbol{R}_{s,n_s-1}^{-1} \boldsymbol{\rho}_{s,n} - \rho_{q_i s_{n_s}}$. The recursive formulation of the matrix of correlations is then $\boldsymbol{\Gamma}_{n_s} = \boldsymbol{\Gamma}_{n_s-1} + \rho^{-1}_{s,n_s|n_s-1}\boldsymbol{\phi}_{n_s}\boldsymbol{\phi}^\top_{n_s}$ and it follows immediately that the difference $\boldsymbol{\Gamma}_{n_s} - \boldsymbol{\Gamma}_{n_s-1} = \rho^{-1}_{s,n_s|n_s-1}\boldsymbol{\phi}_{n_s}\boldsymbol{\phi}^\top_{n_s}$ is positive semi-definite. Finally, this means that the second moment matrix of conditional expectations has the recursive formulation

$$
\begin{aligned}
\boldsymbol{\Omega}_{\zeta,n_s} &= \boldsymbol{\mu}_q\boldsymbol{\mu}_q^\top + \boldsymbol{D}_{\sigma_q}\boldsymbol{\Gamma}_{n_s-1}\boldsymbol{D}_{\sigma_q} + \rho^{-1}_{s,n_s|n_s-1}\boldsymbol{D}_{\sigma_q}\boldsymbol{\phi}_{n_s}\boldsymbol{\phi}^\top_{n_s}\boldsymbol{D}_{\sigma_q} \\
&= \boldsymbol{\Omega}_{\zeta,n_s-1} + \rho^{-1}_{s,n_s|n_s-1}\boldsymbol{D}_{\sigma_q}\boldsymbol{\phi}_{n_s}\boldsymbol{\phi}^\top_{n_s}\boldsymbol{D}_{\sigma_q},
\end{aligned}
\tag{4}
$$

where the last term also can be written as an outer product

$$
\boldsymbol{D}_{\sigma_q}\boldsymbol{\phi}_{n_s}\boldsymbol{\phi}^\top_{n_s}\boldsymbol{D}_{\sigma_q} = \left(\boldsymbol{\sigma}_q \odot \boldsymbol{\phi}_{n_s}\right)\left(\boldsymbol{\sigma}_q \odot \boldsymbol{\phi}_{n_s}\right)^\top.
$$

Proposition 2 (Cost of complexity vs simplicity)**.** *Signs of the increments in the minimized bias-variance trade-off objective and the conditional variance under the true model based on including a signal are*

$$
\begin{aligned}
\chi_{n_s} - \chi_{n_s-1} &= k_{\sigma 0}\mathbf{1}^\top\left\{\boldsymbol{X}^{-1}_{n_s} - \boldsymbol{X}^{-1}_{n_s-1}\right\}\mathbf{1} \geq 0, \\
Var[y|\boldsymbol{\beta}, \boldsymbol{s}_{n_s}] - Var[y|\boldsymbol{\beta}, \boldsymbol{s}_{I,n_s-1}] &= \boldsymbol{\beta}^\top\boldsymbol{D}_{\sigma_q}(\boldsymbol{\Gamma}_{n_s-1} - \boldsymbol{\Gamma}_{n_s})\boldsymbol{D}_{\sigma_q}\boldsymbol{\beta} \leq 0.
\end{aligned}
$$

Proof. The second inequality follows from the observation made in the main-text that $\boldsymbol{\Gamma}_{n_s} - \boldsymbol{\Gamma}_{n_s-1}$ is positive semi-definite. By properties of symmetric positive definite matrices the difference $\boldsymbol{X}^{-1}_{n_s} - \boldsymbol{X}^{-1}_{n_s-1}$ is positive semi-definite if $\boldsymbol{X}_{n_s-1} - \boldsymbol{X}_{n_s}$ is positive semi-definite. This is shown to be the case in Appendix A.4. □

An illustration of Proposition 2 can be found in Figure 2, which covers both the case where the cost of complexity dominates and the case where the benefit of reducing variance under the true model is greater. In Appendix A.5 and Appendix A.6, I show how to extend this analysis to an arbitrary group of additional signals. By induction, the results in Proposition 2 must hold for groups of signals, but I demonstrate that a convenient form similar to 4 exist for a group of signals and, indeed, confirm Proposition 2 for this more general case.

Intuitively what Proposition 2 shows is that adding another signal always (weakly) increases the cost of complexity and lowers the conditional variance under the true model. This is key for interpreting χ as the cost of complexity because it increases when the model is expanded along new dimensions. It is also worth noticing, that the cost of complexity only increases in signals that are fundamentally informative about the factors or the already included signals. This follows from the recursive formulation of the second moment matrix in equation (4) and the definition of the vector $\boldsymbol{\phi}_{1,n_s}$. For a signal uncorrelated with factors $\boldsymbol{q}$ and already included signals $\boldsymbol{s}_{I,n_s-1}$, the vector vector $\boldsymbol{\psi}_{n_s}$ is zero and there is no difference between second moment matrix $\boldsymbol{\Omega}_{\zeta,n_s}$ and $\boldsymbol{\Omega}_{\zeta,n_s-1}$, which propagates to $\boldsymbol{X}_{n_s}$ and $\boldsymbol{X}_{n_s-1}$. This aligns well with the cost of simplicity being the conditional variance under the true model, which clearly is not decreased by conditioning on an irrelevant signal.

3. Heterogeneous agents

In this section, I formulate an extension of the model on the basis of the work-horse asymmetric information model of Grossman and Stiglitz (1980). I make two adjustments to uninformed inference and demand that do not change the classic model but make a difference when informed investors solve the inference problem described in the previous Section 2.1.

Two homogeneous groups of investors, informed and uninformed denoted by $i \in \{I, U\}$, trade a risky asset with independent mean-zero stochastic supply z optimizing demand δ_i over the utility of ultimate profit (equivalent to final wealth, See Appendix B.1). A risk-free asset, which acts as a numeraire with a price and pay-off of one, is available in perfectly elastic supply. Investors are price-takers trading in demand-schedules akin to posting limit orders rather than market orders, see Kyle (1989).

3.1. Uninformed inference

Investors have common priors, which would lead all investors to make the same predictions if endowed with the same information. Investors of type i have information set $\mathcal{F}_i$ and a linear demand function of the form $\delta_i = \psi_i(\hat{y}_i - p)$ where $\hat{y}_i = E[y|\mathcal{F}_i]$ and, with uncertainty aversion α_i, $\psi_i = \{\alpha_i E[(y - \hat{y}_i)^2]\}^{-1}$. In Appendix B.3 and Appendix B.4, I present two foundations for this demand function, respectively a robust profit maximization objective, and CARA-utility with ambiguity aversion. For simplicity, assume that investors know the unconditional mean squared error.[5] In equilibrium, the market clears and the uninformed can extract the signal $s_U := p - \psi_I^{-1}\delta_U = \hat{y}_I - \psi_I^{-1}z$, where ψ_I is the scaling factor of informed demand. I assume that the uninformed investors' prediction of informed investors' prediction is the best linear approximation which I signify by adding a tilde to the expectation, i.e. $\tilde{E}[\cdot]$. It is given by the projection

$$\hat{y}_U = \tilde{E}[\hat{y}_I|s_U] = (1 - \lambda_U)E[\hat{y}_I] + \lambda_U s_U \quad \text{where} \quad \lambda_U = \frac{Var[\hat{y}_I]}{Var[\hat{y}_I] + \psi_I^{-2}\sigma_z^2} < 1, \;\; \sigma_z^2 = Var[z]$$

[5]This assumption can be understood as investors having access to methods such as simulation and cross-validation, to approximate the unconditional mean squared error, and that these methods are accurate enough to abstract away this approximation step and model the approximation as the true unconditional mean squared error. As such, what is abstracted away is the noise in the approximation step.

The best linear approximation minimizes the mean squared error $E[(\hat{y}_I - \hat{y}_U)^2]$, which is consistent with the way the informed chose their predictor (see Section 2.1). In the baseline model where factor loadings are known (equivalent to Grossman and Stiglitz (1980)) this prediction corresponds to the expectation of the pay-off given s_U (see Appendix B.2). The conditional expectation is linear and, therefore, the best linear approximation is *the* best approximation. The formulation here allows for the informed investors noisy estimation of factor loadings that does not depend on a full specification of the distribution of noise. This is true in the baseline model with known factor loadings as well.

3.2. Uninformed mean squared error

The mean squared error of the uninformed investors is a convex combination of the mean squared error of the informed investors and the sum of the unconditional variance of the pay-off and the square of biases scaled by factor means

$$E[(y - \hat{y}_U)^2] = (1 - \lambda_U)\left\{Var[y] + (\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q)^2\right\} + \lambda_U E[(y - \hat{y}_I)^2].$$

Other things equal, a high bias compared to the total cost of complexity χ, which is an element of the informed mean squared error, tends to make the mean squared error of the uninformed higher than that of the informed. Meanwhile, the variance of the pay-off contribute to both both terms but in the informed mean squared error it is through the conditional variance. The conditional variance decomposes into $Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I] = Var[y] - Var[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]]$. Compared to the cost of complexity, signals more informative under the true model, captured by the variance of the expectation with known factors, also tend to make the predictions of the informed investors better than the uninformed. In a baseline model with known factor loadings there is no cost of complexity and the informed mean squared error is always (weakly) lower than the uninformed. I formalize this in Proposition 3.

Proposition 3 (Informed predictions do not always outperform)**.** *The necessary and sufficient condition for out-performance of uninformed predictions by informed predictions is that sum of the bias squared and the variance of the conditional expectation of the pay-off under the objective measure is greater than the cost of complexity*

$$\begin{aligned} E[(y - \hat{y}_U)^2] > E[(y - \hat{y}_I)^2] &\iff \left\{Var[y] + (\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q)^2\right\} > E[(y - \hat{y}_I)^2] \\ &\iff Var[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]] + (\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q)^2 > \chi. \end{aligned}$$

Proof. See Appendix B.5. □

For investors who make their best effort to produce the best prediction possible, the condition can be considered a requirement for anyone to choose to be informed. Notice that the variance of the conditional expectation is the reduction in variance achieved by using a vector of signals under the true model, formally $Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I] = Var[y] - Var[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]]$. The condition echoes the cost and benefits analysis of adding signals in Section 2.4 in that, modulo the squared bias term, the quality of the signals under the true model must be greater than the cost of complexity.

3.3. Price

I close the model extension by deriving the equilibrium price. With linear demand, and an uninformed prediction that is a convex combination, price is a convex combination as well

$$p = (1-\lambda_p)E[\hat{y}_I] + \lambda_p s_U \quad s.t. \quad \lambda_p = \frac{\psi_I + \psi_U \lambda_U}{\psi_I + \psi_U}.$$

Irrespective of the details of the predictions and demand scaling factors ψ_i, the weight on the signal is greater than in the uninformed prediction since

$$\lambda_p > \lambda_U \iff \psi_I + \psi_U \lambda_U > (\psi_I + \psi_U)\lambda_U \iff 1 > \lambda_U.$$

The functional form of price is the same as the expectation of the uninformed investors. If price is viewed as the market's prediction of the risky pay-off, the uninformed are less responsive to the information and supply shocks of s_U than the market since the market also reflect the positioning of informed investors and noise traders. This follows from the asymmetric information and form of demand rather than the inference problem.

4. Predictions

In this section, I highlight a number of predictions where the model (with an without the heterogeneous agents extension) deviate from the baseline model with known factor loadings. To support the analytical analysis, I perform a numerical analysis. For tractability, the numerical analysis is carried out in the minimal setting of two factors, two established signals, and two new signals. For reference, parameters can be found in Table 1, the central matrices in Table 2, key moments in Table 3, and market structure variables in Table 4. First, I impose more structure on the addition of new or more data and discuss what it means for the value of data.

4.1. Value of data

The specification of the inference problem in Section 2.1 introduces two types of data. The signals about factors and the data that estimates of the factor loadings are based on. Throughout the paper, I have lumped in the second type with the estimation technology, because they are both abstracted into the properties of the estimator, and are mathematically summarized by the parameters k_c and $k_{\sigma 0}$. Conceptually, improving estimation technology could mean getting better input data, and better data could mean bigger data. From this perspective, the value of data is a matter of assumption, if it increases k_c^2 it lowers the cost of complexity, if it increases $k_{\sigma 0}$ it raises the cost of complexity. The type of data that affects inference under the true model is the signals. By affecting both the cost of complexity and cost of simplicity the model puts more structure on this data, and its effects on equilibrium outcomes can be compared directly to the baseline model with perfect inference. I extend the study of the inclusion of a group of discrete signals as in Section 2.4 (or groups of signals in Appendix A.5), to the continuous case by analyzing a degree formulation of the problem where the second-moment matrix is given by

$$\boldsymbol{\Omega}_\zeta = \boldsymbol{\Omega}_{\zeta 0} + k_S \boldsymbol{D}_{\sigma_q} \boldsymbol{\Phi}_{n_s} \boldsymbol{R}^{-1}_{s,n_s+|n_s-} \boldsymbol{\Phi}^\top_{n_s} \boldsymbol{D}_{\sigma_q} := \boldsymbol{\Omega}_{\zeta 0} + k_S \boldsymbol{S}, \tag{5}$$

and the formulation of the additional signals matrix is taken from Appendix A.5. If the additional signals are independent of each other and the extant signals and k_S is an integer, it can be interpreted directly as a count of the number of identical signals. More generally, k_S scales the signal group up or down without changing in-between correlations or correlations with the base signals in $\boldsymbol{\Omega}_{\zeta 0}$, providing a way to have more or less of the information it represents. A limitation of this approach is that there is no built-in restriction on k_S that guarantees that the overall correlation structure is feasible. In applications, k_S must be kept at levels that do not generate impossibilities like a negative conditional variance under the true model. However, with this restriction in place, comparative statics with the parameters k_c^2, $k_{\sigma 0}$, and k_S is a useful exercise that captures different aspects of the model.

It is possible to say a bit more about the specification in (5) because the conditional variance under the true model is linear in k_S and the cost of complexity can be shown to be a rational function with an oblique asymptote (i.e. the asymptote is linear in k_S), see Appendix D.1. As the cost of complexity is increasing in k_S and the cost of simplicity decreasing by Proposition 2, the overall mean squared error is eventually linear in k_S, and the tendency comes down to comparing

$$\frac{\partial Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I]}{\partial k_S} = -\boldsymbol{\beta}^\top \boldsymbol{S}\boldsymbol{\beta}, \text{ to } \lim_{k_S \to \infty} \frac{\partial \chi(k_S)}{\partial k_S} = k_{\sigma 0}^2 \mathbf{1}^\top (k_c^2 \boldsymbol{S}^{-1} + \boldsymbol{D}_S^{-1})^{-1}\mathbf{1}.$$

Because $\chi(k_S)$ is everywhere increasing, it approaches the asymptote from below and it follows that $\frac{\partial \chi(k_S)}{\partial k_S}$ is decreasing. Over a given range of k_S, it is, therefore, possible for the mean squared error $E[(y-\hat{y}_I)^2]$ to take one of three shapes. If the derivative of the asymptote dominates the cost of simplicity, it is monotonically increasing. When the reverse is true, the mean squared error can be hump-shaped as the $\frac{\partial \chi(k_S)}{\partial k_S}$ is decreasing or monotonically decreasing. One shape the specification cannot generate is a U-shape where more data is initially beneficial and then eventually becomes a liability.[6] Treating the full second-moment matrix in (5) as one, it is possible that an initial reduction in mean squared error $\boldsymbol{\Omega}_{\zeta 0}$ is gradually undone by a higher k_S until it is eventually better to ignore the full vector of combined signals if it is not possible to separate it. This way the specification can simulate what is in Dugast and Foucault (2020) described as the needle in a haystack problem of big data that is that it becomes harder to find the good signals when there are many to search through.

4.2. Bias and volatility

In addition to covering the optimal bias and volatility across the technology parameters k_c^2 and $k_{\sigma 0}$ as well as the new data parameter k_S, Figure 1 includes a graphical representation of the restriction in Proposition 3 that informed predictions outperform uninformed predictions. For baseline levels $k_c^2 = 1$ and $k_S = 0.5$ the cut-off is at a bit above $k_{\sigma 0} = 0.6$, which is chosen as a harder estimation baseline compared to $k_{\sigma 0} = 0.3$. In Figure 4 the significance of these two levels of difficulty is demonstrated. Under the easy estimation, a stronger new data source input (a higher k_S) results in an overall lower mean squared error of the informed investors' predictor. In contrast, the cost of complexity dominates for the harder estimation problem. There is a basic tension between the constraint of Proposition 3 and predictive deterioration with a stronger new data signal because the former requires the cost of complexity to be bounded and the latter requires it to rise faster than conditional variance falls under the true model. For the special case of a diagonal second moment matrix $\boldsymbol{\Omega}_\zeta = \boldsymbol{D}_{\Omega_\zeta}$, which implies zero mean factors $\boldsymbol{\mu}_q = \mathbf{0}$ such that $\boldsymbol{D}_{\Omega_\zeta} = \boldsymbol{D}_{\Sigma_\zeta}$, and symmetric

[6] By letting correlations decrease in k_S, it is possible to ensure that the lower bound of zero for conditional variance is respected, and with heterogeneity, in the effect, such a U-shape may appear, but in doing so one sacrifices tractability, and I leave the investigation of this extension for future work.

true factor loadings $\boldsymbol{\beta} = \bar{\beta}\mathbf{1}$ the juxtaposition is particularly clear since the constraint is

$$\begin{aligned} Var\left[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]\right] + (\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q)^2 = \bar{\beta}^2 \mathbf{1}^\top \boldsymbol{D}_{\Sigma_\zeta} \mathbf{1} + 0 > k_{\sigma 0}^2 (1 + k_c^2)^{-1} \mathbf{1}^\top \boldsymbol{D}_{\Sigma_\zeta} \mathbf{1} = \chi \\ \iff \bar{\beta}^2 (1 + k_c^2) > k_{\sigma 0}^2, \end{aligned}$$

and the asymptotic condition for a positive derivative of the mean squared error with respect to new data parameter k_S, see Appendix D.1,

$$\begin{aligned} \lim_{k_S \to \infty} \frac{\partial \chi}{\partial k_S} = k_{\sigma 0}^2 (1 + k_c^2)^{-1} \mathbf{1}^\top \boldsymbol{D}_S \mathbf{1} > \bar{\beta}^2 \mathbf{1}^\top \boldsymbol{D}_S \mathbf{1} = \frac{\partial Var\left[y|\boldsymbol{\beta}, \boldsymbol{s}_I\right]}{\partial k_S} \\ \iff k_{\sigma 0}^2 > \bar{\beta}^2 (1 + k_c^2). \end{aligned}$$

While the system of inequalities is only guaranteed to represent a contradiction asymptotically, numerical analysis suggests the intuition that a richer factor structure is necessary to accommodate these conflicting forces. Similarly to panel 1(d) in Figure 1, the restriction of Proposition 3 can be applied to ranges of technology parameter k_c^2 and new data parameter k_S. In the low difficulty case $k_{\sigma 0} = 0.3$ it does not restrict positive values of the two, whereas k_S is limited to be below 0.9 and k_c^2 above 0.8 for high difficulty $k_{\sigma 0} = 0.6$ and plots are adjusted accordingly.

In the first of the three remaining panels of Figure 1, panel 1(a), it is possible to see how initial increases in technology parameter k_c^2 introduces a bias of the estimator while lowering its volatility before eventually decreasing both. Despite the asymmetry of the chosen factor structure, the difference between factors is negligible compared to the difference between moments across both technology k_c^2 and difficulty $k_{\sigma 0}$ in panel 1(c). The comparative statics across new data parameter k_S in panel 1(b) represents more heterogeneity but also a striking symmetry whereby the bias and volatility of each factor visually mirrors one another. Due to the apparent mirroring, it is not obvious from this plot that the cost of complexity is increasing in k_S as stated in Proposition 2, however, the plots in Figure 2 show that it is indeed the case.

4.3. Return predictability

It is the introduction of bias that creates the ex-post predictability of realized price changes r, and bias is optimally chosen by investors to improve the precisions of their predictions. This is in contrast to models with parameter uncertainty (Lewellen and Shanken, 2002; Martin and Nagel, 2021), where bias is with respect to a rational expectations baseline where parameters of the model are known, while those parameters are random variables to agents who have to learn about their realization. Distinction betweeen in-sample and out-of-sample correspond to econometricians testing under the objective measure versus testing under the measure that investors use which I denote by $\boldsymbol{c}^*$, in a reference to the solution to the bias-variance minimization problem of equation (2). Econometricians make predictions through linear projections, which are tantamount to regressions, but estimation includes an estimator choice as in Section 2.1.

In the symmetric agents model of Section 2, my results are only the same as Martin and Nagel (2021), i.e. anomalies seem to exist in-sample, but disappear out of sample, when the quality of estimation technology, parametrized by k_c^2, that investors and econometrician use is the same, or econometrician match the optimal bias and active information set of investors, choice of data sources to include and ignore. By varying k_c^2 it is possible to analyse probable scenarios where econometricians do out of sampling testing, but have access to superior information estimation technology.[7] In this case, a different bias is optimal and I show in Section 4.3.1 that this is a source of predictability. Furthermore, due to the difference in cost of complexity, data sources that are freely available might optimally be ignored by investors using inferior estimation technology but used by econometricians.

The heterogeneous agents model of Section 3 introduces a second source of bias through the learning from prices. It arises from the weight uninformed investors put on their prior.[8] A specific concern for the combination of these two sources of bias is that the out-of-sample testing required to correct the former does not correct the latter. Therefore, econometricians might find an echo of in-sample results in out-of-sample tests.

4.3.1. Returns

Returns can be decomposed into three components that relate to, respectively, the two groups of investors and stochastic supply

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

In the representative agent model of Section 2.1, only the second term remains (with a coefficient of $\lambda_p = 1$) since $p = \hat{y}$ (see Section 2). The inclusion of uninformed investors and stochastic supply introduces the first and last term. As a result, price responsiveness drops below one $\lambda_p < 1$ because the uninformed investors put some weight on their prior. There is no predictability in the supply term and stochastic supply is only relevant in the presence

[7] The objective measure corresponds to the limit $k_c^2 \to \infty$, the best possible technology, see Appendix A.8.

[8] Conceptually, econometricians cannot mimic the beliefs of a representative agent, which is a weighted average of the informed and uninformed beliefs (Biais et al., 2010).

of uninformed investors through its effect on price responsiveness λ_p.

The predictability of the two investor terms can be analyzed by endowing econometricians with an estimation technology corresponding to optimal controls $\boldsymbol{c}_e^*$ that select the vector of signals $\boldsymbol{s}_e$. True out-of-sampling testing is the special case of same estimator $\boldsymbol{c}_e^* = \boldsymbol{c}^*$ and dataset $\boldsymbol{s}_e = \boldsymbol{s}_I$. Problematic in-sample testing corresponds to same dataset but evaluation under the objective measure. If they are available, in-sample testing will include a broader set of signals than the original estimation done by investors as the objective measure is the limiting measure in terms of estimation technology with zero cost of complexity (see A.8).

For ease of exposition, I focus on the case where signals employed by investors are a subset of the econometricians' signals $\boldsymbol{s}_I \subseteq \boldsymbol{s}_e$, and the additional signals in $\boldsymbol{s}_e$, if any, are uncorrelated with $\boldsymbol{s}_I$. I denote the vector of additional signals $\tilde{\boldsymbol{s}}_e$. This restriction is not consequential for *whether* econometricians find predictability or not, but helps to disentangle *where* it comes from.

I reserve the notation $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\varepsilon}_\beta$ for the mean and bias of investors' estimator, e.g. $E[\hat{\boldsymbol{\beta}}|\boldsymbol{c}^*] = E[\boldsymbol{\beta}|\boldsymbol{c}^*] = \boldsymbol{\mu}_\beta$, and notice that in the dataset that econometricians work with these are constants. I denote econometricians' bias by $\boldsymbol{\varepsilon}_{\beta e}$. Finally, I assume that econometricians estimate a cross-sectional average, eliminating the variability in their estimate of factor loadings and, as such, its covariance with investors estimate $\hat{\boldsymbol{\beta}}$, which is in effect evaluated at its mean so $E[\hat{\boldsymbol{\beta}}|\boldsymbol{c}_e^*] = \boldsymbol{\mu}_\beta$.[9]

Proposition 4 (Predictability in returns)**.** *The contribution of the informed component is*

$$E[y - \hat{y}_I|\boldsymbol{c}_e^*, \boldsymbol{s}_e] = \boldsymbol{\beta}^\top(E[\boldsymbol{q}|\boldsymbol{s}_e] - E[\boldsymbol{\zeta}|\boldsymbol{s}_e]) + \boldsymbol{\varepsilon}_\beta^\top E[\boldsymbol{\zeta}|\boldsymbol{s}_e] - \boldsymbol{\varepsilon}_{\beta e}^\top E[\boldsymbol{q}|\boldsymbol{s}_e] \tag{6}$$

$$= (\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\Lambda}_{\tilde{e}}(\tilde{\boldsymbol{s}}_e - \boldsymbol{\mu}_{\tilde{e}}) + (\boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\zeta}, \text{ where } \boldsymbol{\Lambda}_{\tilde{e}} = \boldsymbol{\Sigma}_{q\tilde{s}_e}\boldsymbol{\Sigma}_{\tilde{s}_e}^{-1} \tag{7}$$

and the contribution of the uninformed component is

$$\begin{aligned} E[y - E[\hat{y}_I]|\boldsymbol{c}_e^*, \boldsymbol{s}_e] &= (\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e})^\top (E[\boldsymbol{q}|\boldsymbol{s}_e] - \boldsymbol{\mu}_q) + (\boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\mu}_q \\ &= (\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e})^\top \{\boldsymbol{\Lambda}_I(\boldsymbol{s}_I - \boldsymbol{\mu}_I) + \boldsymbol{\Lambda}_{\tilde{e}}(\tilde{\boldsymbol{s}}_e - \boldsymbol{\mu}_{\tilde{e}})\} + (\boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\mu}_q, \\ &\quad \textit{where} \quad \boldsymbol{\Lambda}_I = \boldsymbol{\Sigma}_{qs_I}\boldsymbol{\Sigma}_{s_I}^{-1} \end{aligned}$$

Proof. The explicit expectations in equation (6) does not depend on $\boldsymbol{c}_e^*$ due to Assumption 2, estimator noise independence from factors and signals, and the cross-sectional mean assumption implies $E[\hat{\boldsymbol{\beta}}|\boldsymbol{c}_e^*] = \boldsymbol{\beta} - \boldsymbol{\varepsilon}_\beta$. Equation (7) uses the independence between $\tilde{\boldsymbol{s}}_e$ and

[9] Alternatively, the exercise can be viewed as an analysis of the unconditional expected coefficients on signals.

$\boldsymbol{s}_I$ which yields

$$E[\boldsymbol{q}|\boldsymbol{s}_e] = \boldsymbol{\mu}_q + \boldsymbol{\Lambda}_I(\boldsymbol{s}_I - \boldsymbol{\mu}_I) + \boldsymbol{\Lambda}_e(\tilde{\boldsymbol{s}}_e - \boldsymbol{\mu}_q) = \boldsymbol{\zeta} + \boldsymbol{\Lambda}_{\tilde{e}}(\tilde{\boldsymbol{s}}_e - \boldsymbol{\mu}_q).$$

The derivation of the uninformed component follows from re-arranging terms after substituting in $E[y|\boldsymbol{c}^*] = \boldsymbol{\mu}_\beta^\top \boldsymbol{\mu}_q$ and once again applying the assumption of independence of $\tilde{\boldsymbol{s}}_e$ and $\boldsymbol{s}_I$ in the second line. □

In contrast to Martin and Nagel (2021), out-of-sample estimation will still be biased if econometricians have access to better technology, especially if that technology leads them to use signals that were available, but too complex to be beneficial at the time of investment. In Figure 4, the estimation quality of the econometricians is increasing and eventually lead to introduction of two additional signals. This affects not only their own predictive coefficients but also those of the actually used signals even though the two are independent. The latter effect is driven by the change in econometricians' own bias, which shifts discretely away from that of investors, visible in the break in the curve of Figure 4(c).

The coefficients on the vector of additional signals is the same for the informed and uninformed component, $\partial E[r|\boldsymbol{c}_e^*, \boldsymbol{s}_e]/\partial \tilde{\boldsymbol{s}}_e = (\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\Lambda}_{\tilde{e}}$, so they are unaffected by the price responsiveness λ_p introduced by the presence of uninformed investor. It is also unaffected by investors bias because these signals are ignored. Therefore, the coefficients on these unused signals are relatively large in absolute terms across the two scenarios, positive or negative true factor loadings, in Figure 4. Assuming variables are properly demeaned, the constant term of the projection is the difference in biases scaled by factor means $E[r|\boldsymbol{c}_e^*, \boldsymbol{s}_e]|_{\boldsymbol{s}_e = E[\boldsymbol{s}_e]} = (\boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\mu}_q$. The coefficients on the used signals also depends on the difference in biases

$$\frac{\partial E[r|\boldsymbol{c}_e^*, \boldsymbol{s}_e]}{\partial \boldsymbol{s}_I} = \{\lambda_p(\boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e}) + (1-\lambda_p)(\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e})\}^\top \boldsymbol{\Lambda}_I = \{\lambda_p \boldsymbol{\varepsilon}_\beta + (1-\lambda_p)\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e}\}^\top \boldsymbol{\Lambda}_I, \tag{8}$$

so when the technological gap between investors and econometricians is not too large, coefficients on used signals and the constant are smaller, because biases are similar. In the heterogeneous agent model, the true factor loadings in the coefficients on used signals are scaled down by a factor of $(1 - \lambda_p)$ compared to the unused signals' coefficients. The representative agent model is at the extreme end of this scaling with the true factor loading term equal to zero.

Only the coefficient on used signals are sensitive to the level of price responsiveness, and as such marks the difference between the representative agent model and the heterogeneous agent model. On the one hand, this means that even if it is possible to control for estimation technology and match investors bias, these coefficients are never zero as would be the null hypothesis in most empirical work and they are sensitive to the out-of-sample echo mentioned above. On the other hand, it is possible to distinguish them from unused signals and other components of returns, such as risk premia, through their sensitivity to variations in price responsiveness.

4.3.2. Return predictability and noise trading

I focus on the level of noise trading σ_z^2 as a source of variation in price responsiveness since it does not affect predictability through other channels, and can more reasonably be taken as exogenous than share of uninformed investors which would be the closest alternative.[10] The scenario in Figure 3 is the easy estimation problem $k_{\sigma 0}$ where there is no difference in investors' and econometricians' information set and all signals, therefore, are affected by shocks to price responsiveness λ_p. In the parametrization of the numerical analysis, the effect of having only half the amount of noise trading, i.e. $\sigma_z^2 = 0.5$ instead of $\sigma_z^2 = 1$, is to lower price responsiveness from 0.92 to 0.87. While this is a modest decrease it is enough to visualize the effect that can also be read of equation (8), which is to shift coefficients by $(\boldsymbol{\beta} - \boldsymbol{\varepsilon}_\beta)^\top \boldsymbol{\Lambda}_I$. Signals that, through the variance-adjusted covariance matrix $\boldsymbol{\Lambda}_I$, load heavily on factors with a difference between factor loading and bias $\beta_i - \varepsilon_{\beta i}$ of the same sign as the signals coefficient will have their coefficients amplified. In markets where more noise trading drives price responsiveness down (which might be considered the natural direction), signals with biases that lead to attenuation of true factor loadings towards zero (including special case $\varepsilon_{\beta i} = 0$) have *larger* coefficients in predictive projections when there is *more* noise trading. This is potentially problematic because noise trading, for good reasons, often is associated with behavioral biases. Even if such behavioral biases effectively generate random noise, a pattern identified in a broad sample may be amplified in a sub-sample that would appear especially representative. It is, however, also a possibility to distinguish between return components as discussed above.

4.4. Price informativeness

The impact of the specifics of investors' inference problem on price informativeness or market efficiency as referred to by Ozsoylev and Walden (2011) has only become more prominent in the context of big data and machine learning on financial markets, see Dugast

[10] While the extension to a variable share of informed versus uninformed investors is straightforward, (see Appendix B.6) allowing the share of informed versus uninformed to vary begs an optimization in the spirit of the original paper to find the optimal share and the introduction of a traditional information acquisition problem, which is well-studied elsewhere and beyond the scope of this paper.

and Foucault (2020) and Farboodi et al. (2020). In the classic formulation of Grossman and Stiglitz (1980), price informativeness is the inverse of the variance of the pay-off conditional on equilibrium price. When price and pay-off are normal random variables, this measure coincides with the mean squared error of a projection, which has led to empirical estimation strategies based on regression analysis, see Dessaint, Foucault, and Frésard (2020). For comparability, I also base my measure of price informativeness on a projection. As discussed in Section 3.1, this projection is also the best linear approximation. Price informativeness is given by

$$E[(y - E[y|p, \boldsymbol{\beta}])^2]^{-1} = \left\{ Var[y] - \frac{\lambda_p^2}{\lambda_p^2} \frac{Cov[y, \hat{y}_I]^2}{Var[s_U]} \right\}^{-1} = \left\{ \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_q \boldsymbol{\beta} - \frac{(\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\mu}_\beta)^2}{Var[\hat{y}_I] + \psi_I^{-2} \sigma_z^2} \right\}^{-1}. \tag{9}$$

Equation (9) shows how price responsiveness λ_p does not affect price informativeness. The effect of uninformed trading on price informativeness is just to scale supply by the square inverse scaling factor of the informed investors, ψ_I^{-2}, while larger supply σ_z^2 unambiguously decreases it. This is true in the baseline model of known factor loadings as well. The difference is in the inflation of variance of the informed predictor compared to the variance of the expectation with known factors, and the attenuation of the covariance

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

Higher estimation noise $\boldsymbol{\sigma}_\beta$ unambiguously inflates the variance of the predictor through the cost of complexity, while the role of the bias is less clear cut. However, numerical analysis supports the tendency of variance inflation and covariance attenuation. The presence of the interaction term $\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta$ in both the denominator and numerator implies that if a configuration of bias amplifies rather than attenuates covariance it will with at an even higher inflation of variance.

Taking noisy inference as a given, the unattainable base line is less interesting than the fact that price informativeness under this condition is subject to a trade-off between bias and variance just as the predictor of (informed) investors. Figure 6 is based on, respectively, the equilibrium inverse price informativeness (mean squared error), and the counterfactual of a social planner optimizing price informativeness, i.e. solving the non-linear problem of maximizing equation 9 with respect to controls $\boldsymbol{c}$ that control the bias and variance of the estimator $\hat{\boldsymbol{\beta}}$ (see Section 2.2). Carried out under the objective measure, a stronger new data source signal (higher k_S) raises the price informativeness (lowers the mean squared error), even though the baseline difficulty is high $k_{\sigma 0} = 0.6$, which means that the mean squared error of the informed predictor is increasing (see Figure 2). Of the three dimensions considered, only sophisticated estimation technology (high k_c^2) leads to conversion between the private optimization and the planner's optimization. While the two other parameters are restricted by the requirement that informed predictions outperform (Proposition 3), the tendency to convergence happens in the unrestricted direction towards zero, but convergence is not reached before the hard cut-off at zero. A planner with an ability to invest in improving technology $\uparrow k_c^2$, lowering the baseline difficulty $\downarrow k_{\sigma 0}$ or making new data sources available $\uparrow k_S$, would find all beneficial but might prefer the first as it aligns the goals of private individuals with its own. Especially since the incorporation of new data sources might require better technology to be attractive to private individuals, as it is possible for informed investors' predictive power to be deteriorating even while price informativeness is improving.

4.5. Excess price volatility

Another measure of market quality is price volatility and in particularly in excess of the volatility of dividends (Shiller, 1980). In the model, this comparison correspond to the contrast between price variance

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

and the variane of the pay-off $Var[y]$. In the baseline model with known factor loadings, conditional expectation of the pay-off is $Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I] = Var[y] - Var[E[|y|\boldsymbol{\beta}, \boldsymbol{s}_I]]$, and $Var[E[|y|\boldsymbol{\beta}, \boldsymbol{s}_I]]$ replaces $Var[\hat{y}_I]$ in $Var[p]$. As such, excess price variance is not possible in the representative agent model. In the heterogeneous agent model it requires large amounts of noise trading and the off-setting effect of uninformed traders correcting for it to be small[11] (with no uninformed demand, price responsiveness is unity $\lambda_p = 1$). The additional noise from the estimation process makes excess price variance more prevalent in the parameter space and it can occur even in the representative agent model of Section 2, i.e. $Var[\hat{y}_I] > Var[y]$ is possible.

[11] The effect of more noise trading on λ_p has two counter-acting forces. More noise leads the uninformed to put more weight on their prior ($\downarrow \lambda_U$), which lowers λ_p, but they also take smaller positions ($\downarrow \psi_U$), which increases λ_p as informed investors and noise traders make up a bigger share of the market. Conversely, when uninformed investors form their best estimate by copying informed investors as best as they can (see Section 3.1) they do not adjust the same way for estimation noise as noise trading. Because higher estimation noise goes into $Var[\hat{y}_I]$, they will actually *increase* the weight they put on their signal derived from price ($\uparrow \lambda_U$).

4.6. Short-term price reversals

There is a long tradition of using short-term price reversals to disentangle informed trading from liquidity demand, see Hendershott and Menkveld (2014), since the price pressure from liquidity demand can be expected to be reversed in the spirit of Campbell et al. (1993). Predicting price change by price gets at short term price reversals in the context of a two period model assuming a constant before-price of $E[p]$, see Breon-Drish (2015). I first illustrate how the price pressure of liquidity demand relates to price reversals before adding realized estimated factor loadings $\hat{\boldsymbol{\beta}}$ as an additional conditioning variable. I once again apply projection as the linear approximation of the expectation and in, addition to price, I condition on negative stochastic supply z (liquidity demand)

$$\tilde{E}[r|p,z] = \boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q + \frac{Cov[y,\hat{y}_I] - \lambda_p Var[\hat{y}_I]}{\lambda_p Var[\hat{y}_I]}(p - E[p]) - \frac{Cov[y,\hat{y}_I]}{\lambda_p Var[\hat{y}_I]}\psi_I^{-1} \underbrace{(-z)}_{\substack{\text{liquidity}\\ \text{demand}}},$$

(derivations that can be found in Appendix C.1). The non-zero constant of the projection $\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q$ generates a tendency for price drift that might over time be picked up as momentum or long term reversal. This component is analysed in Section 4.3 on return predictability and here I focus on short-term reversals instead.

It is useful to consider that in the baseline model without noisy inference, covariance of the pay-off and the informed predictor is the variance of the predictor $Cov[y, E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]] = Var[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]]$. Therefore, the coefficients simplify to $(1-\lambda_p)/\lambda_p$ and $-\psi_I^{-1}/\lambda_p$ respectively, and positive supply shock tends to be followed by negative return. Directionally, the impact of a large supply shock is the same as in the baseline model with known factor loadings if the covariance between informed prediction and pay-off is positive. It is still positive but it is attenuated if $Cov[y,\hat{y}_I] < Var[\hat{y}_I]$, which is likely as a sufficient condition is for this inequality is $Var[y] < Var[\hat{y}_I]$.

More important for interpreting empirical results on reversals, conditioning on the realization of the factor loadings in the projections yields identical coefficients on rescaled negative supply $-\psi_I^{-1}z$ and factor-mean weighted realized factor loadings $\boldsymbol{\mu}^\top \hat{\boldsymbol{\beta}}$.

Proposition 5 (Short-term reversals)**.** *Both liquidity demand* $(-z)$ *and realized estimated factor loadings* $\hat{\boldsymbol{\beta}}$ *generate expected short-term reversals with a common factor in marginal* *effects*

$$\tilde{E}\left[r|p,z,\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q + \frac{Cov[y,\hat{y}_I] - \lambda_p(Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q)}{\lambda_p(Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q}(p - E[p])$$
$$- \frac{Cov[y,\hat{y}_I]}{\lambda_p(Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q)}(-\psi_I^{-1} z) - \frac{Cov[y,\hat{y}_I]}{\lambda_p(Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q)}\boldsymbol{\mu}_q^\top(\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta).$$

Proof. See Appendix C.1.1. □

While it is unreasonable to expect econometricians to directly observe either stochastic supply z or factor loadings estimates $\hat{\boldsymbol{\beta}}$, the significance of Proposition 5 is that any variable that correlates with price reversals could be correlated with liquidity demand, but could equally well be correlated with noisy estimation. Whether it is meaningful to group these two types of noise into one will depend on the context. In contrast to the results in Section 4.3, which were largely driven by the bias of the noise inference the effect for short-term reversals is a function of the variance of the estimation process.

Short term reversals can also be linked to price variance (see Section 4.5) since higher estimation noise tends to attenuate coefficients on both the price, supply shocks, and realized factor loadings per Preposition 5. Higher trading noise only affects coefficients through price responsiveness λ_p which will tend to decrease as uninformed investors put more weight on their prior. Analysing price variance and price reversals is, therefore, useful for understanding whether a likely common factor is: estimation noise (attenuated coefficients), or trading noise (amplified coefficients).

In the limiting case of the estimation technology parameter k_c^2 growing large presented in Figure 7, both of these relations exist, and there is additionally a negative relation between the coefficient on price and price variance when driven by estimation noise and a positive one when driven by trading noise. For the parametrization of this numerical analysis, the relations between trading noise, price variance, and reversal coefficients are consistent across different levels of technological sophistication for the case of attenuation bias captured by $+\bar{\beta}$. The case of amplifying bias, $-\bar{\beta}$, is more complicated because the relation between the base-parameter of estimation difficulty $k_{\sigma 0}$ and price variance changes over the range of the estimation technology parameter k_c^2, from being positive for lower values to eventually being negative and some instability in between. The interesting aspect of the limiting case is how the relationship between price variance and the reversal coefficients again prevails despite the changes in the relationship with the base parameter with amplification bias, i.e. price variance is decreasing in $k_{\sigma 0}$.

It is not the only one possible, but from an empirical perspective the most straightforward identifying assumption is that estimation bias is dominated by attenuation bias, in which case the opposing relations between price variance and the price reversal coefficient emerges. It is additionally attractive because of its significance in turning noise trading into predictability in returns as covered in Section 4.3.2.

4.6.1. Trading volume

With several explanations for returns, a natural additional dimension of market data to consider is trading volume following Campbell et al. (1993). Realized trading volume v is given by

$$\begin{aligned} v &= \frac{1}{2}\left\{|\delta_I| + |\delta_U| + |z|\right\} = \frac{1}{2}\left\{|\delta_I| + |-\delta_U| + |z|\right\} \\ &= \frac{\psi_I}{2}\left\{ \left|(1-\lambda_p)(\hat{y}_I - E[\hat{y}_I]) - \lambda_p \psi_I^{-1} z\right| + (1-\lambda_p)\left|(\hat{y}_I - E[\hat{y}_I]) + \psi_I^{-1} z\right| + |\psi_I^{-1} z| \right\}, \end{aligned}$$

see Appendix C.2. From this expression it is possible to point out a difference between liquidity demand and noisy estimation. Other things equal, the impact of an absolute increase in stochastic supply/liquidity demand $|z| = |-z|$ is always positive for non-zero demand and supply since

$$\frac{\partial v}{\partial |z|} = \frac{1}{2}\psi_I^{-1}\left\{-\lambda_p \operatorname{sign}(\delta_I)\operatorname{sign}(z) - (1-\lambda_p)\operatorname{sign}(\delta_U)\operatorname{sign}(z) + 1\right\} \geq 0.$$

Conversely, an informed prediction with larger absolute deviation from its expected value is ambiguous

$$\frac{\partial v}{\partial |\hat{y}_I - E[\hat{y}_I]|} = (1-\lambda_p)\operatorname{sign}(\hat{y}_I - E[\hat{y}_I])\left\{\operatorname{sign}(\delta_I) - \operatorname{sign}(\delta_U)\right\},$$

however, for large enough deviations the effect will be positive as the trading between informed and uninformed investors dominates the trading flow. The classic intuition that conditioning on large trading volume with low expected returns helps to identify liquidity demand, following Campbell et al. (1993), breaks down in the presence of noisy inference.

4.7. Fund performance

By identifying informed investors as sophisticated funds who invest in inference technology and the data it requires and uninformed investors as their simpler counterparts, I compare performance as measured by their expected profits. Profit of investors under a measure $\boldsymbol{c}_j^*$ is derived in Appendix C.3.1,

$$\begin{aligned}
E[\pi_I] &= \psi_I \left\{(1-\lambda_p)\left(Cov[y,\hat{y}_I|\boldsymbol{c}_j^*] - \lambda_p Var[\hat{y}_I]\right) + \lambda_p^2 \psi_I^{-2}\sigma_z^2\right\}, \text{ and} \\
E[\pi_U] &= \psi_U \left\{(\lambda_p-\lambda_U)\left(\lambda_p Var[\hat{y}_I] - Cov[y,\hat{y}_I|\boldsymbol{c}_j^*]\right) + (\lambda_p-\lambda_U)\lambda_p \psi_I^{-2}\sigma_z^2\right\} \\
&= \psi_I(1-\lambda_p)\left\{\left(\lambda_p Var[\hat{y}_I] - Cov[y,\hat{y}_I|\boldsymbol{c}_j^*]\right) + \lambda_p \psi_I^{-2}\sigma_z^2\right\},
\end{aligned}$$

see Appendix C.3.1. Contrasting ex-ante expected profits under investors' measure $\boldsymbol{c}^*$ with the large numbers average realized profit (expectation under objective measure or equivalent limiting measure $\boldsymbol{c}_\infty^*$) comes down to comparing $Cov[y,\hat{y}_I|\boldsymbol{c}^*]$ and $Cov[y,\hat{y}_I]$. The difference between the two defines the ex-post surprise investors are expected to experience under the model. In Appendix C.3.3, I show that there is no ex-post surprises with respect to the total profits of the investor base and the surprise is a transfer between informed and uninformed investors equal to

$$E[\pi_I] - E[\pi_I|\boldsymbol{c}^*] = -(E[\pi_U] - E[\pi_U|\boldsymbol{c}^*]) = \psi_I(1-\lambda_p)(Cov[y,\hat{y}_I] - Cov[y,\hat{y}_I|\boldsymbol{c}^*]).$$

The condition for a surprise ex-post out-performance of informed investors, and by symmetry the under-performance of uninformed investors, is whether the covariance between the pay-off and the informed predictor is over- or underestimated under the contemporary measure compared to the objective measure. In Appendix C.3.2, I link expected profit under the contemporaneous measure, i.e. expected out-performance by informed investors, to the condition for informed investors making better predictions presented in Proposition 3. Under the parametrization of the numerical analysis, Figure 8 establishes a clear relationship between this condition and the sign of $\bar{\beta}$, and Figure 9 shows that it carries over to profits. With attenuation bias ($+\bar{\beta}$), the out-of-sample surprise is in favor of informed investors, (sophisticated investors, for some period perhaps quantitative funds) and with amplification bias, the effect is in the direction of hype followed by disappointment. Linking these conditions to Section 4.3.1 and Section 4.6 the common environment is characterized by the shared condition of attenuation bias, and fund performance is a way to identify periods where this condition is likely true (under the model). The period leading up to a period of out-performance by quantitative funds (or another identifier of informationally sophisticated funds) is a good candidate for the joint analysis of return predictability and short-term price reversals.

5. Predictability of the equity risk premium

In this section, I focus on a specific case of predictability in the asset pricing literature that the noisy estimation with changing technology provides an explanation for: the predictability of the equity risk premium, specifically with respect to the predictive variables surveyed by Welch and Goyal (2008) and the literature inspired hereby (Campbell and Thompson, 2008; Rapach et al., 2010; Neely et al., 2014; Buncic and Tischhauser, 2017; Hammerschmid and Lohre, 2018). Variation in estimation technology offers an explanation for two empirical patterns found within and across these studies: vanishing predictability over time (in the data) but stronger predictability overall between studies (see Section 5.1). As a single time series with variation through predictive variables rather than a cross-section of assets the application is better suited of the narrower representative agent model of Section 2 than the extended heterogeneous agent model of Section 3. Therefore, in Section 6, I provide some considerations for further empirical work collecting and extending observations made in Section 4 and Section 4.1.

5.1. Predicting the equity risk premium

An empirical pattern found in Welch and Goyal (2008) and confirmed in later studies with other empirical strategies and auxiliary data (e.g. Buncic and Tischhauser (2017)) is that of, over time, an initial out-performance of the historical mean followed by deteriorating performance[12] driven by the variables identified in the first study. The turn-around falls in the earlier 1990s and as such it follows the rise of the private computer in the 1980s and coincides with the early years of the internet. A second pattern that appears *between* studies is one of empirical approaches in subsequent papers out-performing approaches in earlier papers. These effects have a natural interpretation in terms of the estimation technology of respectively investors and econometricians.

In the model, the parameter k_c^2 can be interpreted as the quality of estimation technology. By providing subscripts I for investors and e for econometricians, the first empirical pattern can be understood as fixing k_{ce}^2 and increasing k_{cI}^2. Within a given study the empirical method used by econometricians is the same through time while the data it is applied to changes. Over time, the estimation technology of investors could reasonably be improving as new techniques become available.

The second empirical fact, can be seen as fixing a path for k_{cI}^2 and increasing k_{ce}^2 where

[12]Dependent on specific set-up this might mean under-performance or performance only on par with that of the historical mean. Introducing a floor of zero for the prediction of the equity risk premium as proposed by Campbell and Thompson (2008), generally helps to avoid under-performance.

later studies represents a better estimation technology than earlier. Abstracting estimation technology into the single dimension of k_c^2 suggests that there is a hierarchy of methods. A priori, it is not obvious how to rank approaches and without assumptions on the data-generating process no such ranking can exists given the no free lunch theorem for learning algorithms (Wolpert, 1996). However, ex posteriori, some empirical strategies should out-perform others if there is something to learn from the data. One point to address, is that out-performance between studies coincides with the introduction of new data. As discussed in Section 4.1, in the model, data that is used for estimation of the factor loadings is reflected in k_c^2 (or $k_{\sigma 0}$, the estimation difficulty). Conceptually, however, allowing this dimension to vary at the same time as fundamental data, the structure imposed by the model is loosened substantially. Therefore, in my replication of these patterns in Figure 12, I do not generate the second pattern by replicating the studies directly. Instead, I fix the data used and shows how regularised linear approaches can outperform plain vanilla ordinary least squares.[13] In the calibration of Section 5.1.1, I focus on the first empirical pattern and investigate which shift in investors estimation technology k_{cI}^2 is required to generate it.

5.1.1. Calibration

In the calibration, I focus on moments that rely only on a subset of parameters. I consider ten of the predictive variables studied by Welch and Goyal (2008) and use the updated time series data from Amit Goyal's website.[14] In addition to the assumption that the representative agent model is a meaningful first-order representation of the phenomenon, I operate under the identifying assumption that investors' signal set is the same over the period. With the discussion of the potency of a mismatch between employed signals and available signals in Section 4.3.1, it is a very relevant alternative hypothesis, and one to keep in mind when considering the results of the calibration. Approaching it directly, however, requires richer data as it introduces more degrees of freedom. The first set of moments is the difference in variance adjusted expected coefficients on the econometricians signals between the two sub-periods

$$\left\{ E\left[\frac{\partial E[r_2|\boldsymbol{c}_e^*, \boldsymbol{s}_e]}{\boldsymbol{\partial} s_e}\right] - E\left[\frac{\partial E[r_1|\boldsymbol{c}_e^*, \boldsymbol{s}_e]}{\partial \boldsymbol{s}_e}\right]\right\} Var[\boldsymbol{s}_e] = (\boldsymbol{\varepsilon}_{\beta 2} - \boldsymbol{\varepsilon}_{\beta 1})^\top \boldsymbol{D}_{\sigma_q} \boldsymbol{R}_{q s_I} \boldsymbol{R}_{s_I}^{-1} \boldsymbol{R}_{s_I s_e}$$

[13] Per the discussion of the impossibility of an a priori ranking of methods it is not generally clear that such an out-performance should exists, but the application is inspired by the uses of regularization in later studies (Rapach et al., 2010; Buncic and Tischhauser, 2017)

[14] The variables are listed in Table 5.

empirically estimated by the averages of the coefficients from rolling regressions with 30 years windows.[15] Notice how keeping the econometricians estimation technique fixed along with the signal set of the investor means that terms involving the bias of the econometricians drop out. The second set, relies on a different assumption to eliminate terms involving the econometricians bias, which is that the benefit of in sample estimation eventually makes the estimation error of econometricians on a given data-set small in comparison to investors. It is the variance adjusted expected coefficients on the econometricians signals over the full sample

$$E\left[\frac{\partial E[r|\boldsymbol{c}_e^*, \boldsymbol{s}_e]}{\boldsymbol{\partial} s_e}\right] Var[\boldsymbol{s}_e] = [w\boldsymbol{\varepsilon}_{\beta 2} + (1-w)\boldsymbol{\varepsilon}_{\beta 1}]^\top \boldsymbol{D}_{\sigma_q} \boldsymbol{R}_{q s_I} \boldsymbol{R}_{s_I}^{-1} \boldsymbol{R}_{s_I s_e}$$

where the weights are approximated by the number of observations in the two sub-periods. In Figure 13, I confirm that the coefficients targeted can generate the empirical pattern of predictive out-performance followed by deteriorating performance when applied to the data. Finally, I include the difference in the unconditional expectation,

$$E[r_2] - E[r_1] = (\boldsymbol{\varepsilon}_{\beta 2} - \boldsymbol{\varepsilon}_{\beta 1})^\top \boldsymbol{\mu}_q,$$

as a target moment.

I fix a number of parameters up front and limit the estimation to the correlation structure and the second period investor estimation technology quality parameter k_{cI2}. I maintain the structure from the numerical analysis of two factors with common mean and variance, $\bar{\mu}_q$ and $\bar{\sigma}_q$, but summarize investors' group of signals in one signal which requires two correlation parameters, ρ_{qs_I1} and ρ_{qs_I2}. Varying $\bar{\mu}_q$, $\bar{\sigma}_q$, the baseline $k_{\sigma 0}$, and first period estimation difficulty k_{cI1} has little impact on the directional effects as long as certain relations are maintained to ensure convergence, $\bar{\mu}_q < \bar{\sigma}_q$ and $k_{\sigma 0} \approx |k_{cI1}|$. I run the estimation iteratively in a two step procedure alternating between minimizing the mean squared error between theoretical and empirical moments over the correlation parameters and then k_{cI2}. I find that then convergence occur, it happens after a few iterations.

I pick $k_{\sigma 0}$ and $\bar{\mu}_q$ to match the baseline values studied in the numerical analysis and calibrate $\bar{\sigma}_q$ such that the correlations between each factor and investors' signal group come out with comparable magnitudes in the estimation, $\rho_{qs_I1} \approx 0.3$ and $\rho_{qs_I2} \approx -0.21$. Apart from these two correlation parameters, I estimate a correlation parameter for each of the

[15] In Figure 10 and Figure 11 I investigate the in-sample fit of expanding and contracting as well as rolling windows. The pattern for shorter windows is consistent with some amount of over-fitting in sample as the *shorter* samples achieve noticeably higher scores but have very poor untabulated out-of-sample performance. The interpretation of in-sample analysis as converging to the objective measure, following Martin and Nagel (2021), is not reasonable for the shorter windows and I focus on 20 and 30 years windows instead.

ten predictive variables,which, from the perspective of the model, are the econometricians signals. Most parameter values can be found in Table 6, but the correlation parameters for the predictive signals I instead visualize in Figure 14, which clearly shows that the magnitude of correlation with investor signals is the largest for the valuation ratios of dividend- and earnings-to-price. Since the estimation is done with respect to variance-covariance corrected measures, should not be taken to mean that these are the only relevant predictive variables. Rather, valuation is estimated to be the strongest channel through which the variables relate to investors signals which is in line with the view that this group of predictive variables represent signals about fundamentals as opposed to, for instance, sentiment.

The key number of interest is the shift in estimation technology. A direct way to look at it is through the percentage increase in the magnitude of parameter k_{cI} which comes out to $k_{cI2}/k_{cI1} - 1 \approx 233\%$. For a more contextual view that also reflects the estimated information structure, the change in investors optimal bias can be calculated based on the calibration. With two factors the vector has two elements which are virtually the same both experiencing a growth of $\varepsilon_{\beta i,2}/\varepsilon_{\beta i,1} - 1 \approx 82\%$. Taking the information structure into account the magnitude of the shift is attenuated substantially. However, what is perhaps more surprising about this estimate is the direction of the shift. The estimated optimal bias has *grown*. Since there is a trade-off between bias and variance increasing bias and decreasing cost of complexity both follow from the increasing quality of estimation technology. For a visual demonstration of this mechanism see Figure 1(a). Recognizing that bias can be optimal helps to explain the empirical pattern.

The shift in estimation technology is substantial, which might to a certain extent reflect the discretization into just two periods. The turnaround in Figure 12 is, however, rather sharp and the alternative hypothesis of technological development also changing the composition of investors signal group is reasonable. Either way, the calibration is consistent with technological development historically playing a large role in predictability of returns.

6. Further directions for empirical work

One of the key challenges in working empirically with the model, is that taking the concern that both investors and econometricians face an estimation problem in forming predictions about pay-offs or returns seriously makes it problematic to estimate key ingredients of the model such as the cost of complexity directly from the data. Extending to the cross-section, separating the estimation of $\hat{\boldsymbol{\beta}}$ and factor expectations $\boldsymbol{\zeta}$ à la Kelly, Pruitt, and Su (2019) or in a non-linear fashion as in Gu et al. (2021) would allow for more targeted contrasts of sources predictability. E.g. once a factor structure is estimated cost of complexity under various signal configuration and parameter choices could be matched to moments of predictability in the cross-section. Alternatively, separating returns into components that more closely match what is pay-off and price in the model, by proxying expected pay-offs by earnings expectation or focusing on returns around earnings-announcement would again allow a more targeted estimation.

Irrespectively of the cross-sectional approach, the heterogeneous agent extension highlights promising ways to sub-sample the data. Most directly related to return predictability, the decomposition of signals available to in Section 4.3 shows how only signals employed by investors are affected by price responsiveness. This suggest an empirical strategy of sub-sampling on proxies for noise trading to sort components of predictability on their sensitivity to this cross-sectional variation.[16] Another prediction related to noise trading is that when higher noise trading leads to lower price responsiveness the coefficient on signals estimated with attenuation bias are larger in sub-samples with more noise (see Section 4.3.2). While it might be natural to assume that attenuation bias is more common than amplification that can generally be an identifying assumption that is hard to provide direct support for. In the context of the model, relations between price variance and short-term price reversals with respect to levels of noise trading is particularly stable under attenuation bias (see Section 4.6), and in the time series, fund performance can be related to the dominant type of bias, as periods of success (better than expected performance) for sophisticated investors are more likely under attenuation bias (see Section 4.7).

7. Conclusion

The complex prediction problems faced by investors in financial markets have a number of implications for equilibrium outcomes. Complexity generates a cost of expanding predictive models of the risky pay-off with new signals. Investors optimally trade off bias for precision and the benefit of including a signal for the associated cost of complexity. Advances in estimation techniques such as machine learning methods mitigate the issues of complexity for investors. In the study of financial markets, these advanced methods, however, require careful application to undo rather than amplify the bias optimally introduced by investors. Empirically, the effect of changing estimation technology on predictability is not only relevant for new methods going forward but and can explain historical patterns of predictability as well. For new methods that likely lowers the cost of complexity, a high level of caution is warranted in assessing their likely future performance through historical back-testing. Analysing historic performance over time rather than summarising performance in an aggre-

[16]These component may be predictive variables or relevant transformations of the data.

gate statistic is a mitigating measure, as well as analysis of cross-sectional sub-samples in various dimensions especially around proxies for noise trading.

References

Admati, A. R., 1985. A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Markets. Econometrica 53, 629.

Bai, J., Philippon, T., Savov, A., 2016. Have financial markets become more informative? Journal of Financial Economics 122, 625–654.

Biais, B., Bossaerts, P., Spatt, C., 2010. Equilibrium Asset Pricing and Portfolio Choice Under Asymmetric Information. Review of Financial Studies 23, 1503–1543.

Breon-Drish, B., 2015. On Existence and Uniqueness of Equilibrium in a Class of Noisy Rational Expectations Models. The Review of Economic Studies 82, 868–921.

Brunnermeier, M. K., Parker, J. A., 2005. Optimal Expectations. The American Economic Review 95, 1092–1118.

Buncic, D., Tischhauser, M., 2017. Macroeconomic factors and equity premium predictability. International Review of Economics & Finance 51, 621–644.

Campbell, J., Grossman, S. J., Wang, J., 1993. Trading volume and serial correlation in stock returns. The Quarterly Journal of Economics 108, 905–939.

Campbell, J. Y., Thompson, S. B., 2008. Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? Review of Financial Studies 21, 1509–1531.

Cochrane, J. H., 2011. Presidential Address: Discount Rates. The Journal of Finance 66, 1047–1108.

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

Didisheim, A., Ke, S., Kelly, B. T., Malamud, S., 2023. Complexity in Factor Pricing Models. SSRN Electronic Journal .

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., 2021. A Growth Model of the Data Economy p. 52.

Gabaix, X., 2014. A Sparsity-Based Model of Bounded Rationality *. The Quarterly Journal of Economics 129, 1661–1710.

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

Gilboa, I., Schmeidler, D., 1989. Maxmin expected utility with non-unique prior. Journal of mathematical economics 18, 141–153.

Goldstein, I., Spatt, C. S., Ye, M., 2021. Big Data in Finance. The Review of Financial Studies 34, 3213–3225.

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.

Hammerschmid, R., Lohre, H., 2018. Regime shifts and stock return predictability. International Review of Economics & Finance 56, 138–160.

Harvey, C. R., 2017. Presidential Address: The Scientific Outlook in Financial Economics: Scientific Outlook in Finance. The Journal of Finance 72, 1399–1440.

Hastie, T., Tibshirani, R., Friedman, J. H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2. New York: springer.

Hellwig, M. F., 1980. On the aggregation of information in competitive markets. Journal of Economic Theory 22, 477–498.

Hendershott, T., Menkveld, A. J., 2014. Price pressures. Journal of Financial Economics 114, 405–423.

Hou, K., Xue, C., Zhang, L., 2020. Replicating Anomalies. The Review of Financial Studies 33.

Jensen, T. I., Kelly, B. T., Pedersen, L. H., 2021. Is there a replication crisis in finance? NBER .

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

Kelly, B. T., Pruitt, S., Su, Y., 2017. Instrumented Principal Component Analysis. SSRN Electronic Journal .

Kelly, B. T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134, 501–524.

Kyle, A. S., 1985. Continuous Auctions and Insider Trading. Econometrica 53, 1315.

Kyle, A. S., 1989. Informed Speculation with Imperfect Competition. The Review of Economic Studies 56, 317.

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 .

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

Molavi, P., Tahbaz-Salehi, A., Vedolin, A., 2021. Model Complexity, Expectations, and Asset Prices. NBER p. 44.

Neely, C. J., Rapach, D. E., Tu, J., Zhou, G., 2014. Forecasting the Equity Risk Premium: The Role of Technical Indicators. Management Science 60, 1772–1791.

O'Hara, M., 2003. Presidential Address: Liquidity and Price Discovery. The Journal of Finance 58, 1335–1354.

Ozsoylev, H. N., Walden, J., 2011. Asset pricing in large information networks. Journal of Economic Theory 146, 2252–2280.

Pastor, L., Veronesi, P., 2009. Learning in Financial Markets. Annual Review of Financial Economics 1, 361–381.

Rapach, D. E., Strauss, J. K., Zhou, G., 2010. Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. Review of Financial Studies 23, 821–862.

Shiller, R., 1980. Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? Tech. Rep. w0456, National Bureau of Economic Research, Cambridge, MA.

Sims, C. A., 2003. Implications of rational inattention. Journal of Monetary Economics 50, 665–690.

Van Nieuwerburgh, S., Veldkamp, L., 2010. Information Acquisition and Under-Diversification. Review of Economic Studies 77, 779–805.

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

Wolpert, D. H., 1996. The Lack of A Priori Distinctions Between Learning Algorithms. Neural Computation 8, 1341–1390.

Figures and Tables

Table 1: Baseline parameter values.

	Notation	Value
Estimation technology level	k_c^2	1
Difficulty of estimation problem	$k_{\sigma 0}$	0.3
New data sources strength	k_S	0.5
Common component in factor loadings	$\bar{\beta}$	0.6
Individual components in factor loadings	$\tilde{\boldsymbol{\beta}}$	$(1\ 0.3)^\top$
Common uncertainty aversion	α_I, α_U	1
Intensity of stochastic supply (noise trading)	σ_z^2	1
Common factor mean	$\bar{\mu}_q$	0.2
Common factor volatility	$\bar{\sigma}_q$	1
Common factor scaling factor	$(1/\lvert\mathbf{1}\rvert)\mathbf{1}$	$0.7 \times \mathbf{1}$
Correlation between factors	ρ_q	0
Correlation between established signals	ρ_{s0}	0.2
Correlation between new signals	ρ_{s1}	-0.5
Correlation inc. signals and factors (diagonal matrix)	$\text{diag}(\boldsymbol{R}_{qs0})$	$(0.5\ 0.25)^\top$
Correlation new signals and factors (diagonal matrix)	$\text{diag}(\boldsymbol{R}_{qs1})$	$(0.25\ 0.5)^\top$
Correlation inc. signals and new (independent)	$\boldsymbol{R}_{ss}$	$0\boldsymbol{I}$

Table 2: Baseline factor structure.

	Notation	Value
Factor mean outer product	$\boldsymbol{\mu}_q \boldsymbol{\mu}_q^\top$	$\begin{pmatrix} 0.02 & 0.02 \\ 0.02 & 0.02 \end{pmatrix}$
Factor covariance	$\boldsymbol{\Sigma}_q$	$\begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$
Cond. exp. of factors		
Est. signals covariance term	$\boldsymbol{\Sigma}_{\zeta 0}$	$\begin{pmatrix} 0.13 & -0.01 \\ -0.01 & 0.03 \end{pmatrix}$
Unscaled new signals cov. term	$\boldsymbol{S}$	$\begin{pmatrix} 0.04 & 0.04 \\ 0.04 & 0.17 \end{pmatrix}$
Second moments	$\boldsymbol{\Omega}_\zeta = \boldsymbol{\mu}_q \boldsymbol{\mu}_q^\top + \boldsymbol{\Sigma}_{\zeta 0} + k_S \boldsymbol{S}$	$\begin{pmatrix} 0.17 & 0.03 \\ 0.03 & 0.14 \end{pmatrix}$

Table 3: Baseline moments.

	Notation	Value $+\bar{\beta}$	Value $-\bar{\beta}$
Variance of pay-off	$Var[y]$	0.20	0.20
Cost of complexity	χ	0.02	0.02
Cond. var. of pay-off true model	$Var[y \vert \boldsymbol{\beta}, \boldsymbol{s}_I]$	0.14	0.14
Mean squared error informed	$E[(y - \hat{y}_I)^2]$	0.15	0.15
Mean squared error uninformed	$E[(y - \hat{y}_U)^2]$	0.17	0.16
Var. informed predictor	$Var[\hat{y}_I]$	0.04	0.11
Expectation of inf. pred.	$E[\hat{y}_I]$	0.07	-0.15
Expectation of pay-off true	$E[y \vert \boldsymbol{\beta}]$	0.11	-0.11
Var. inf. pred. true model	$Var[E[\hat{y}_I \vert \boldsymbol{\beta}, \boldsymbol{s}_I]]$	0.06	0.06
Cov. inf. predictor and pay-off			
Contemporary measure	$Cov[y, \hat{y}_I \vert \boldsymbol{c}^*]$	0.03	0.10
Objective measure	$Cov[y, \hat{y}_I]$	0.04	0.08

Table 4: Baseline market structure.

	Notation	Value $+\bar{\beta}$	Value $-\bar{\beta}$
Informed position scaling factor	ψ_I	6.6	6.6
Uninformed position scaling factor	ψ_U	6.0	6.3
Uninformed responsiveness to price-signal	λ_U	0.64	0.82
Price responsiveness to shocks	λ_p	0.82	0.91

Table 5: Predictive variables. In depth variable descriptions can be found in Welch and Goyal (2008).

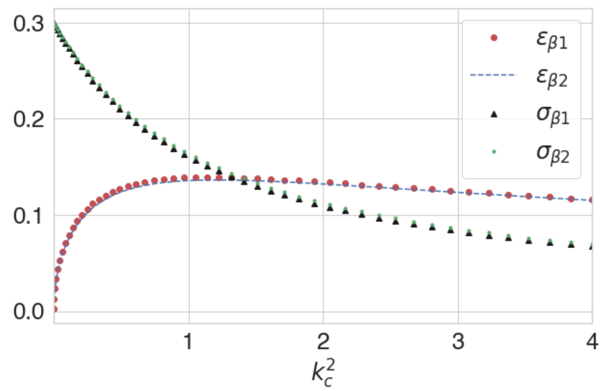
Name	Notation
Dividend price ratio	dp
Earnings price ratio	ep
Stock variance	$svar$
Book to market value	bm
Corporate Issuing Activity	$ntis$
Treasury bills	tbl
Long term yield	lty
Default yield spread	dfy
Default return spread	dfr
Inflation	$infl$

Table 6: Parameters of the calibration (see Section 5.1.1) excluding predictive variables correlation parameters (see Figure 14).

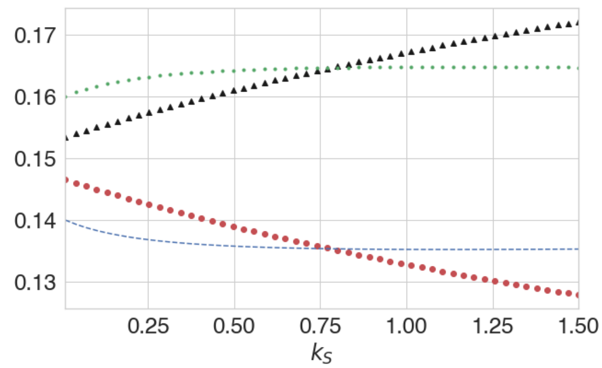
Name	Notation	Value
Weight in full sample average expected coefficients	w	0.64
Common factor mean	$\bar{\mu}_q$	0.2
Common factor volatility	$\bar{\sigma}_q$	0.8
Difficulty of estimation problem	$k_{\sigma 0}$	0.3
Investors' estimation technology level period 1	k_{cI1}	-0.3
Investors' estimation technology level period 2	k_{cI2}	-1.0
Correlation between investors' signal group and factor 1	$\rho_{qs_I 1}$	0.3
Correlation between investors' signal group and factor 2	$\rho_{qs_I 2}$	-0.21

Fig. 1. Bias ε_β and volatility σ_β. Panel 1(d) shows cost of complexity χ and the left hand side of the inequality of Proposition 3 as $\dagger = Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I] + (\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\mu}_q)^2$.

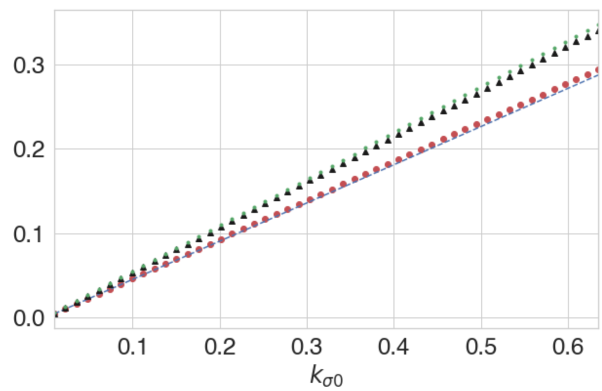
(a) Better estimation technology



(b) Stronger new data source



(c) Harder estimation problem



(d) Informed investors better informed

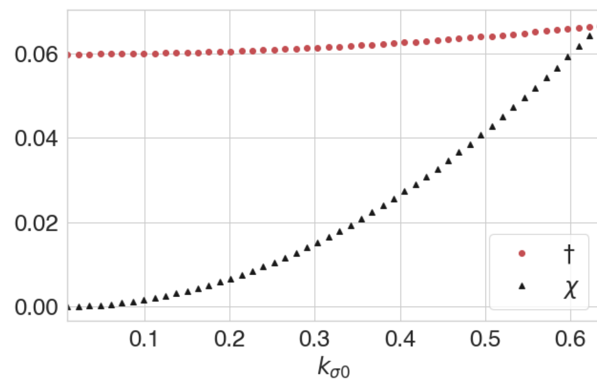
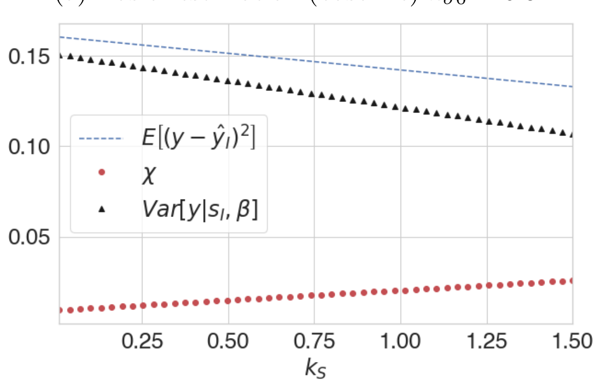


Fig. 2. Mean squared error of the informed predictor $E[(y - \hat{y}_I)^2]$, cost of complexity χ, and conditional variance of the pay-off under the true model $Var[y|\boldsymbol{s}_I, \boldsymbol{\beta}]$. Stronger new source of data (higher k_S) can lead to higher or lower mean squared error. Given the baseline parametrization the two scenarios are captured by high or low estimation difficulty $k_{\sigma 0}$.

(a) Easier estimation (baseline) $k_{\sigma 0} = 0.3$



(b) Harder estimation (baseline) $k_{\sigma 0} = 0.6$

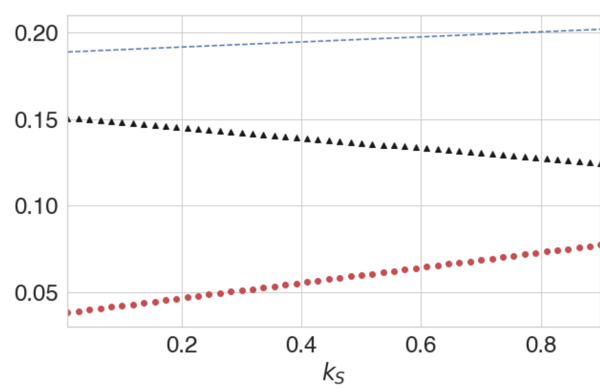


Fig. 3. Easier estimation problem ($k_{\sigma 0} = 0.3$) means that informed investors uses all information. The microstructure fundamental of trading noise mainly shifts coefficient curves to larger absolute values. Price responsiveness λ_p is respectively ≈ 0.92 and ≈ 0.87.

(a) Less noise trading $\sigma_z^2 = 0.5$ (b) More noise trading $\sigma_z^2 = 1$

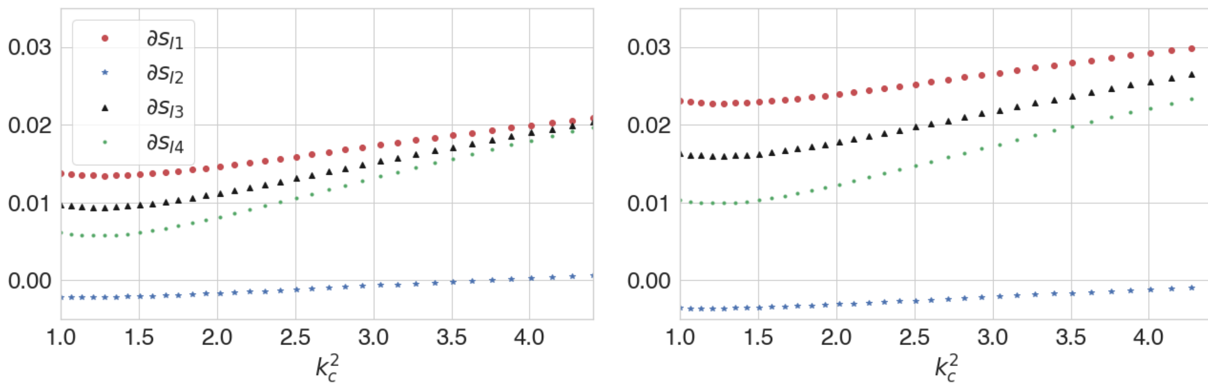
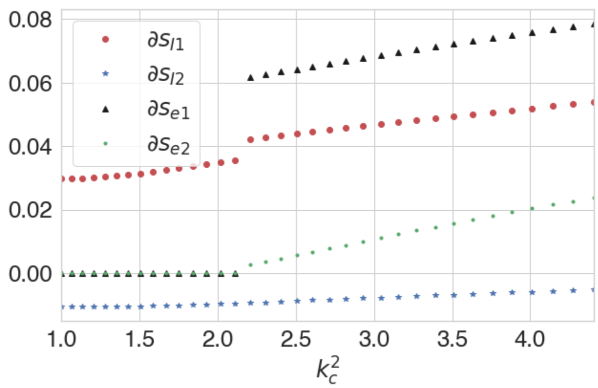
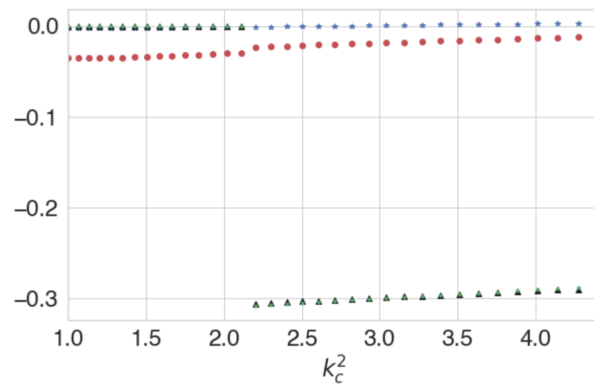


Fig. 4. Harder estimation problem ($k_{\sigma 0} = 0.6$) means that the information set of informed investors and econometricians eventually differs. Noise trading is at the baseline level $\sigma_z^2 = 1$. Information differences demonstrates a notable break. Signals $\boldsymbol{s}_I$ and $\boldsymbol{s}_e$ are independent, but their coefficients are connected through econometricians' bias.

(a) Positive $\bar{\beta}$



(b) Negative $\bar{\beta}$



(c) Econometricians' bias

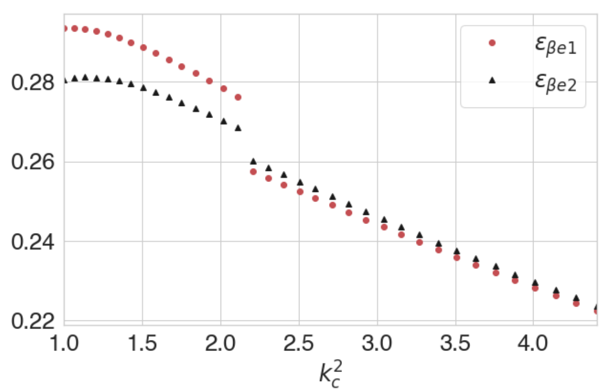
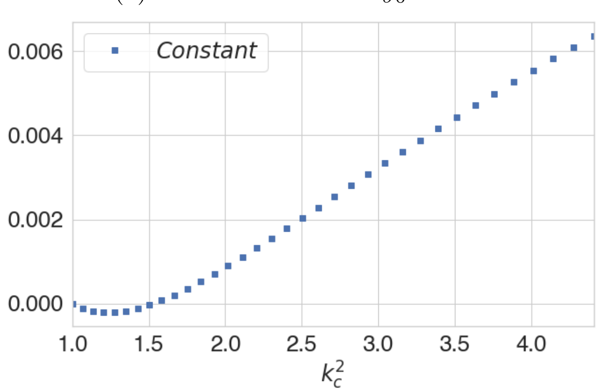


Fig. 5. The bias in the constant term of the projection of price changes r onto signals is increasing in the gap between investors' and econometricians' bias as the latter a decreases toward zero (see Figure 4).

(a) Easier estimation $k_{\sigma 0} = 0.3$



(b) Harder estimation $k_{\sigma 0} = 0.6$

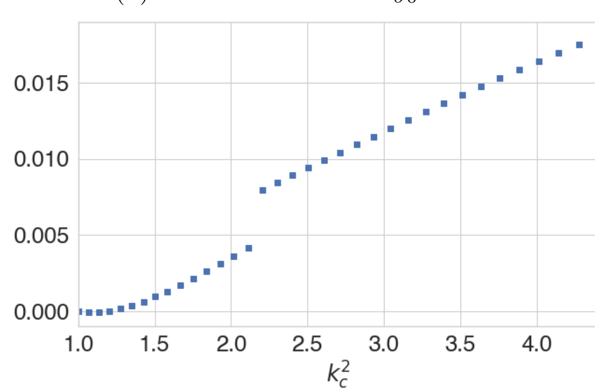


Fig. 6. Mean squared error of prediction based on price contrasting optimal choice of estimator made by the informed investors (private) and a social planner optimizing for price informativeness. All plots under the hard estimation scenario (baseline) $k_{\sigma 0} = 0.6$. Easy estimation scenario is similar in most cases, except for the contrast between same and opposite sign true factor loadings, $+\bar{\beta}$ and $-\bar{\beta}$ respectively, when varying the new data parameter k_S, which is less pronounced with both graphs looking more like the plot in column 6(b).

(a) Positive $\bar{\beta}$ (b) Negative $\bar{\beta}$

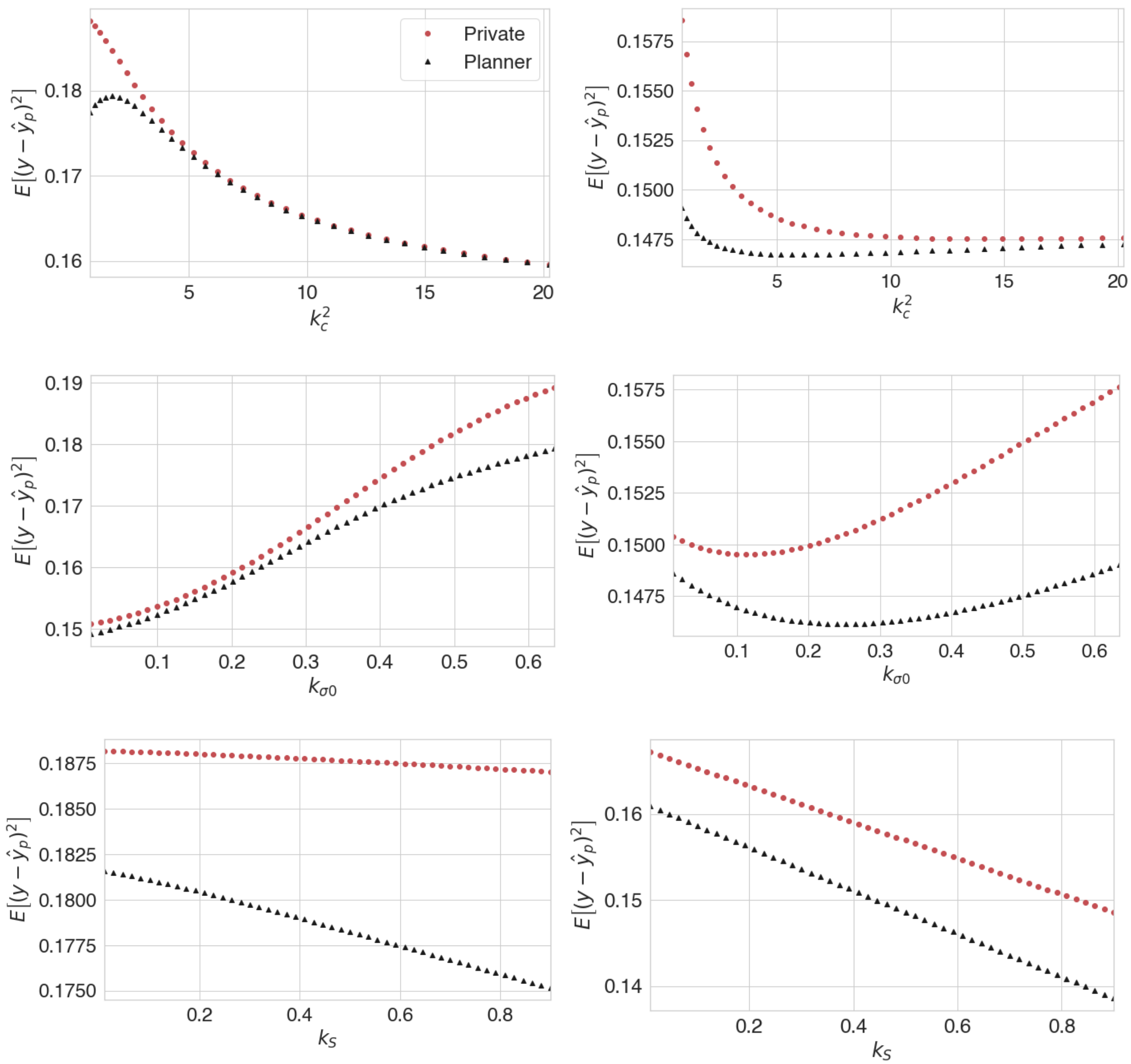
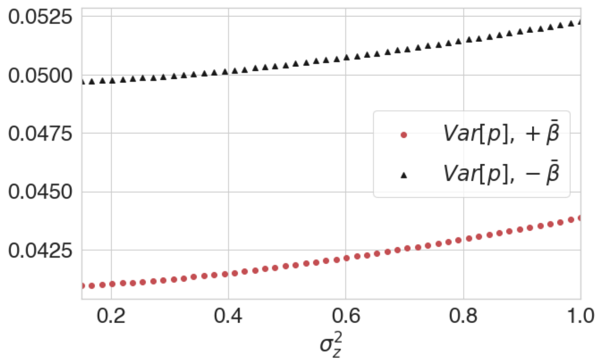
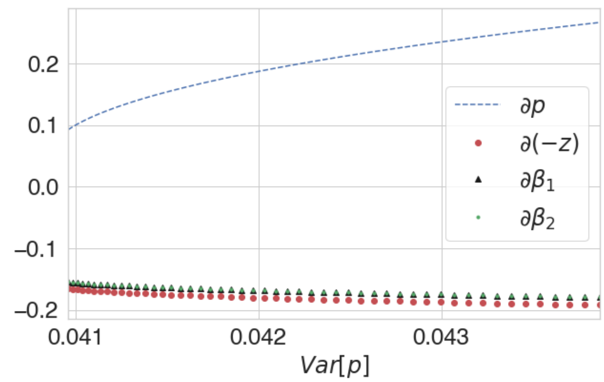


Fig. 7. Variance of price and reversals coefficients with respect to trading noise σ_z^2, and estimation noise driven by estimation difficulty $k_{\sigma 0}$. Price variance increasing in trading noise also holds at lower levels of estimation technology parameter k_c^2, as well as for same sign bias and true factor loadings $(+\bar{\beta})$. Meanwhile price variance decreasing in estimation difficulty $k_{\sigma 0}$ is reversed at lower levels. Overall the tendency for opposite trends in coefficients is found at different levels of technological sophistication in parameter regions allowed by the constraint of Proposition 3 (informed predictions outperform).

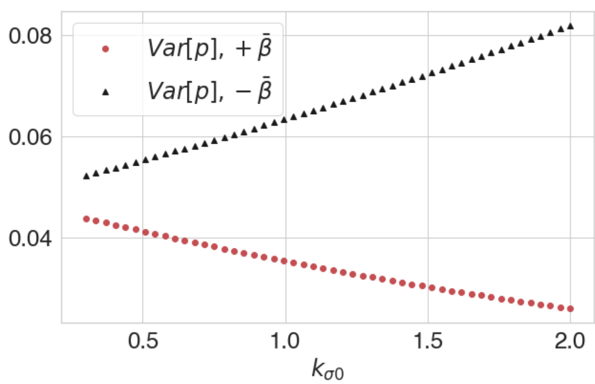
(a) Limiting case $k_c^2 = 100$, trading noise σ_z^2



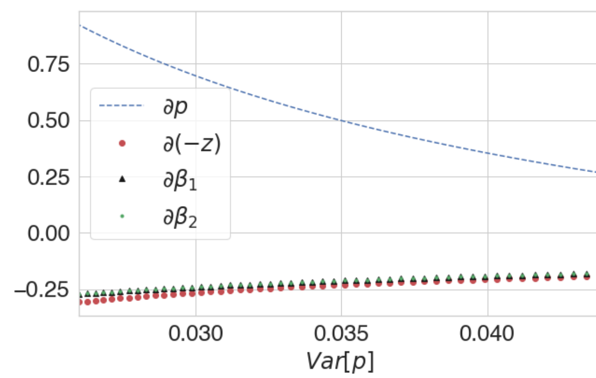
(b) Positive $\bar{\beta}$, trading noise σ_z^2



(c) Limiting case $k_c^2 = 100$, estimation noise $k_{\sigma 0}$



(d) Positive $\bar{\beta}$, estimation noise $k_{\sigma 0}$



(e) Negative $\bar{\beta}$, estimation noise $k_{\sigma 0}$

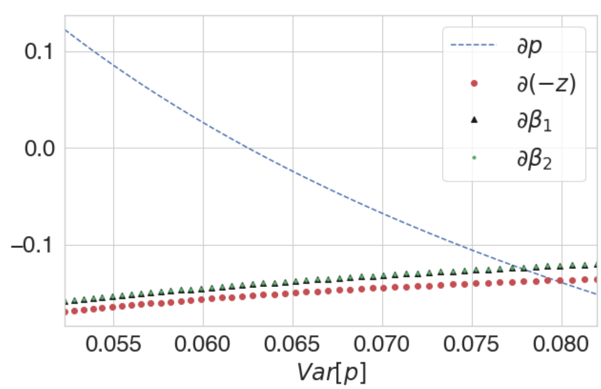


Fig. 8. Variance and covariance of informed predictor and pay-off under contemporary measure c^* and objective measure. Both the shape of the curves as well as the ordering of the covariances is determined by the sign of the sign of true factor loadings parametrized by $\bar{\beta}$. The variation across estimation difficulty is included for comparability with Figure 9.

(a) Easier estimation $k_{\sigma 0} = 0.3$, negative $\bar{\beta}$ (b) Harder estimation $k_{\sigma 0} = 0.6$, positive $\bar{\beta}$

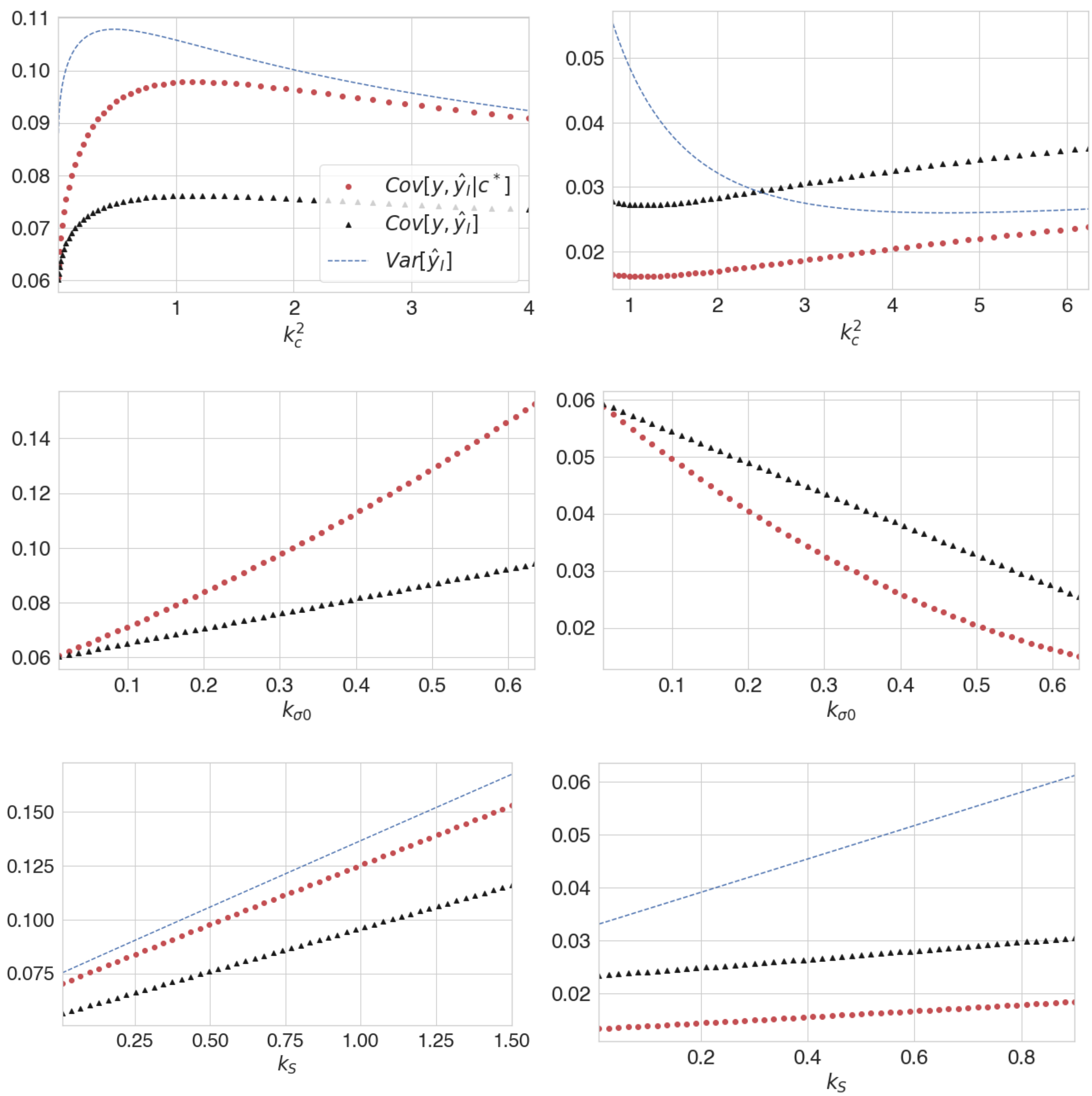


Fig. 9. Fund performance under the contemporary measure and objective measure. The ordering of the corresponding covariances determine whether average expected profit or average realized profit is higher. The baseline difficulty influences the trend of the two other parameters, most notably new data k_S, which increases profits in the hard estimation scenario.

(a) Easier estimation $k_{\sigma 0} = 0.3$, negative $\bar{\beta}$ (b) Harder estimation $k_{\sigma 0} = 0.6$, positive $\bar{\beta}$

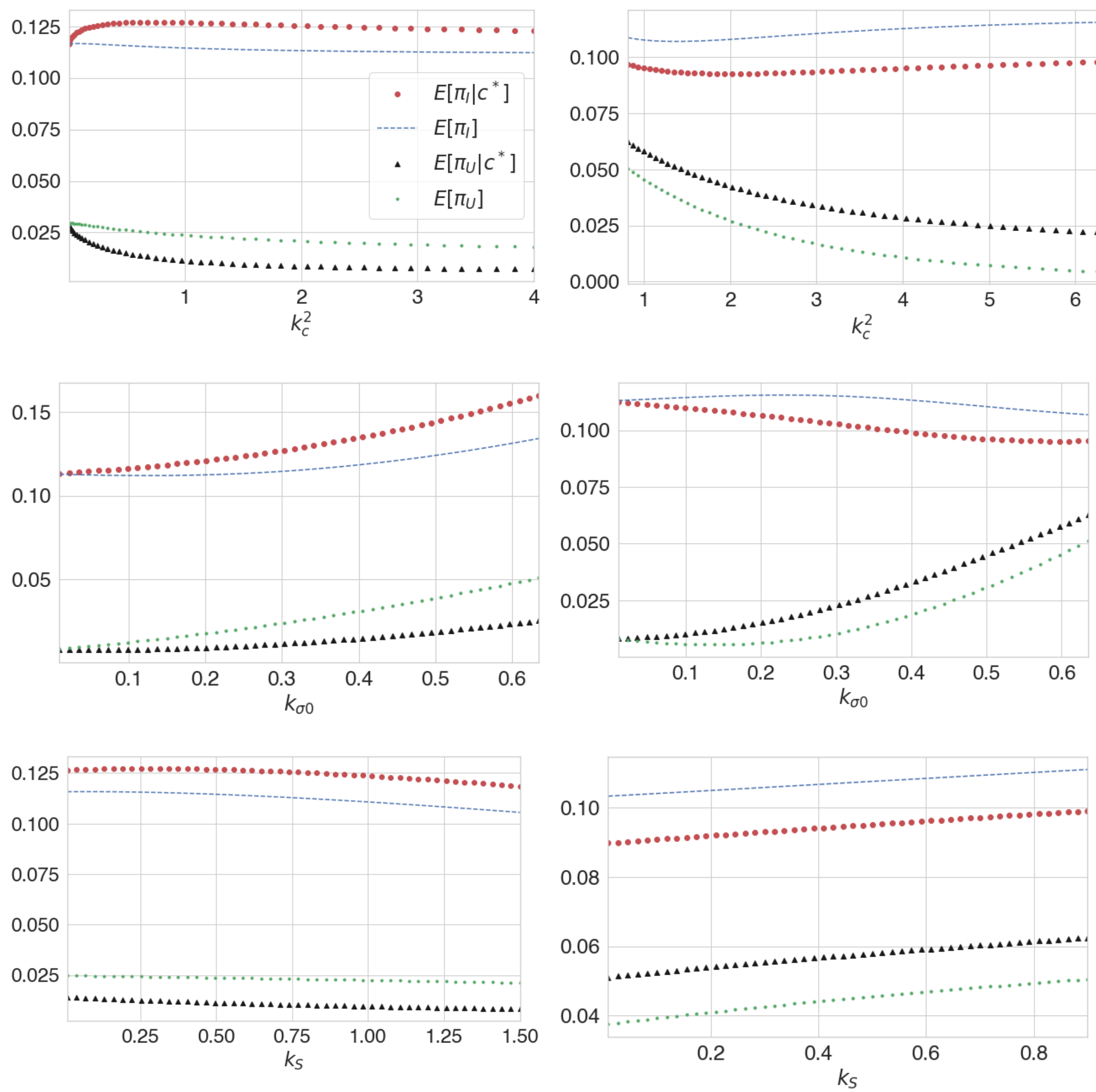


Fig. 10. Expanding and contracting windows adjusted in-sample R-squared. $Adj\ R^2 = 1 + \frac{n+1}{n-p}(R^2 - 1)$ where n is number of observations and p is number of parameters.

(a) Expanding window

(b) Contracting window

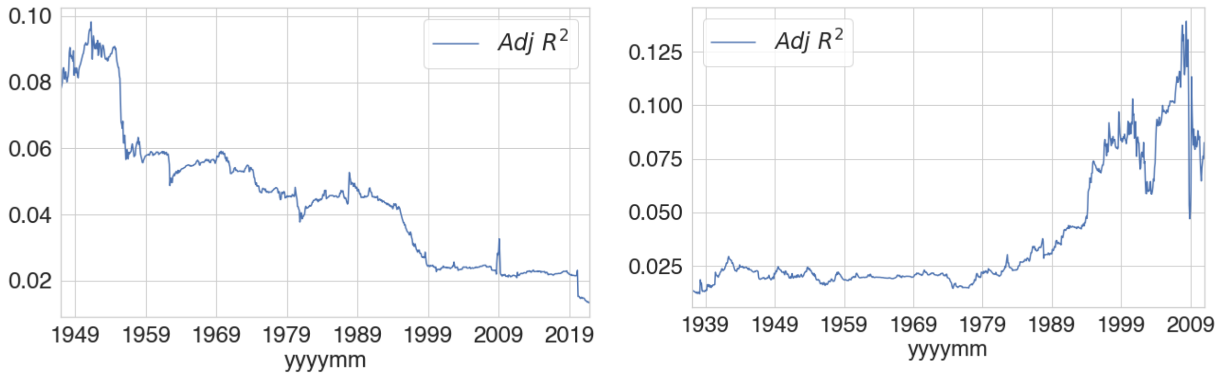


Fig. 11. Rolling windows adjusted in-sample R-squared. $Adj\ R^2 = 1 + \frac{n+1}{n-p}(R^2 - 1)$ where n is number of observations and p is number of parameters. Each curve represent the adjusted R-squared of a model estimated over the number of years shown in the legend. Plots are smoothened by plotting 10 years rolling averages of the adjusted R-squared.

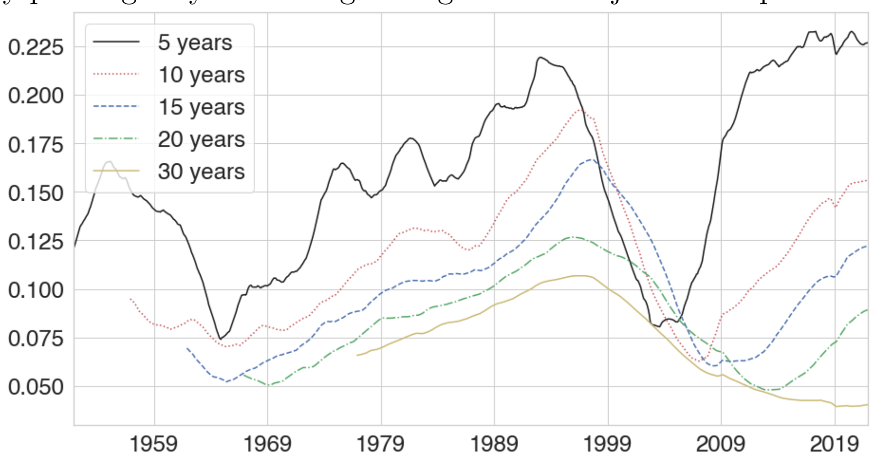
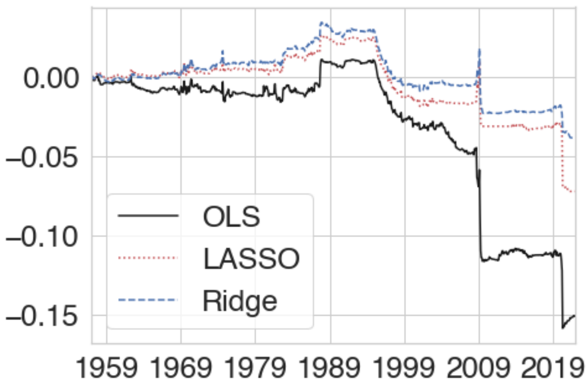
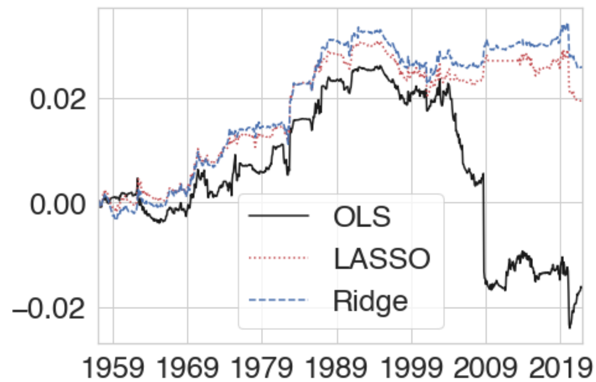


Fig. 12. Predictive performance compared to historical mean of different methods over rolling windows of 20 and 30 years and with and without the Campbell and Thompson (2008) zero floor that floors predictions at zero denoted CT. Except for 30 years with CT, OLS is a clear under-performer. All plots present deteriorating performance in the second sub-period (starting around 1991) and all expect 20 years OLS without CT present out-performance in the first.

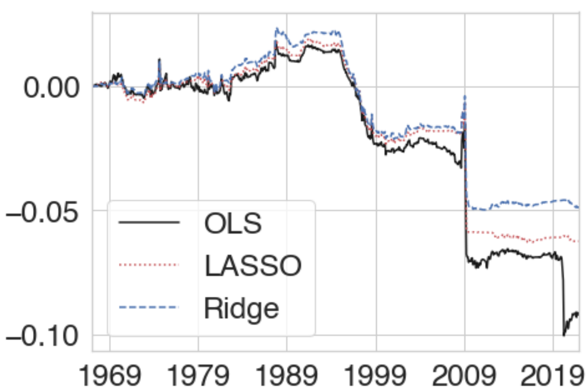
(a) 20 years, no CT



(b) 20 years, CT



(c) 30 years, no CT



(d) 30 years, CT

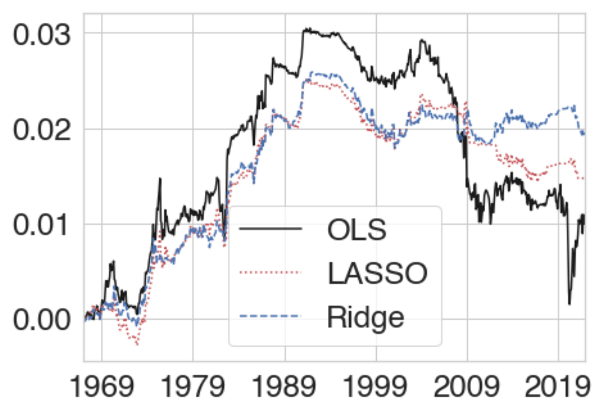


Fig. 13. Predictive performance compared to historical mean of expected coefficient targets for estimation.

(a) Expected coefficients of sub-periods (b) Expected coefficients full period

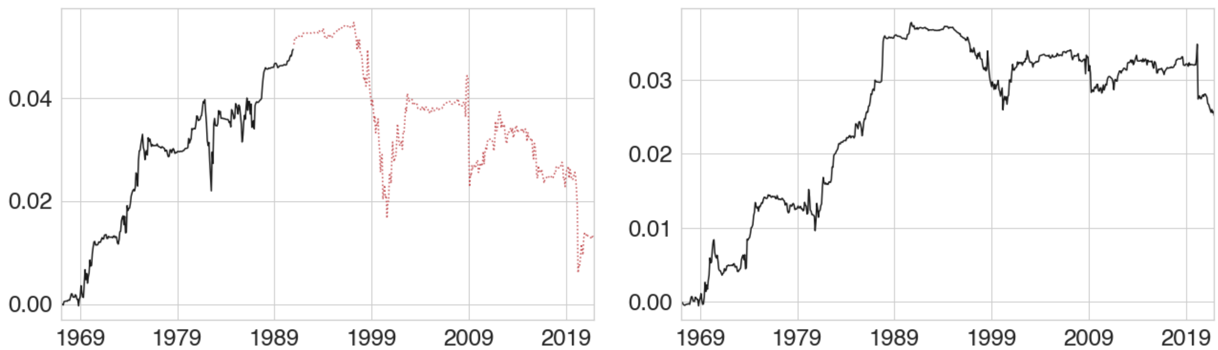
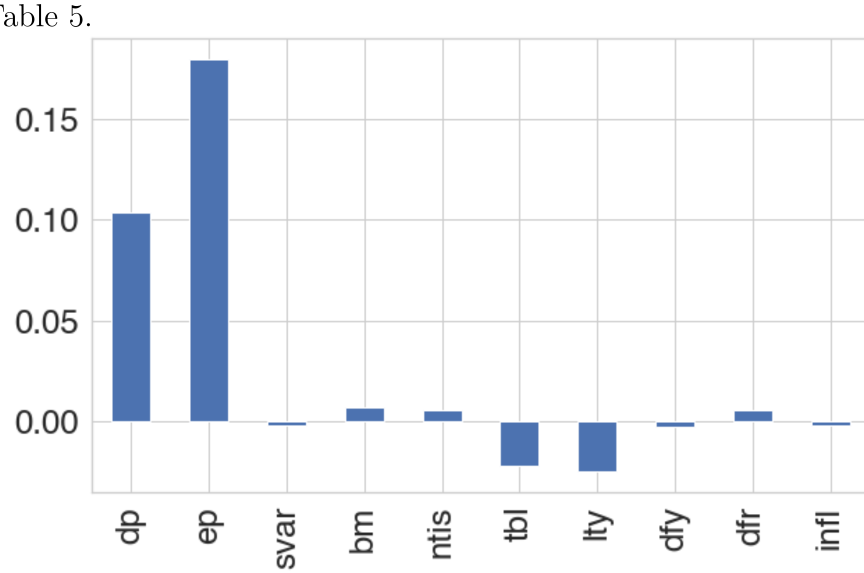


Fig. 14. Predictive variable correlation parameters estimated in calibration. Description of variables in Table 5.



Appendix A. Inference problem

A.1. Variance of dot-product of independent random vectors

For random vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ with mean and variance $\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x \ \forall x \in \{v, w\}$ using the trace operator tr

$$\begin{aligned}
Var\left[\boldsymbol{v}^\top \boldsymbol{w}\right] &= E\left[\left(\boldsymbol{v}^\top \boldsymbol{w}\right)^2\right] - \left(E\left[\boldsymbol{v}^\top \boldsymbol{w}\right]\right)^2 = E\left[\boldsymbol{v}^\top \boldsymbol{w}\boldsymbol{w}^\top \boldsymbol{v}\right] - \left(E\left[\boldsymbol{v}\right]^\top E\left[\boldsymbol{w}\right]\right)^2 \\
&= E\left[\operatorname{tr}\left(\boldsymbol{v}\boldsymbol{v}^\top \boldsymbol{w}\boldsymbol{w}^\top\right)\right] - \left(\boldsymbol{\mu}_v^\top \boldsymbol{\mu}_w\right)^2 = \operatorname{tr}\left(E\left[\boldsymbol{v}\boldsymbol{v}^\top\right] E\left[\boldsymbol{w}\boldsymbol{w}^\top\right]\right) - \left(\boldsymbol{\mu}_v^\top \boldsymbol{\mu}_w\right)^2 \\
&= \operatorname{tr}\left(\left\{\boldsymbol{\mu}_v\boldsymbol{\mu}_v^\top + \boldsymbol{\Sigma}_v\right\}\left\{\boldsymbol{\mu}_w\boldsymbol{\mu}_w^\top + \boldsymbol{\Sigma}_w\right\}\right) - \left(\boldsymbol{\mu}_v^\top \boldsymbol{\mu}_w\right)^2 \\
&= \operatorname{tr}\left(\boldsymbol{\mu}_v\boldsymbol{\mu}_v^\top \boldsymbol{\mu}_w\boldsymbol{\mu}_w^\top\right) + \operatorname{tr}\left(\boldsymbol{\mu}_v\boldsymbol{\mu}_v^\top \boldsymbol{\Sigma}_w\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}_v\boldsymbol{\mu}_w\boldsymbol{\mu}_w^\top\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}_w\boldsymbol{\Sigma}_v\right) - \left(\boldsymbol{\mu}_v^\top \boldsymbol{\mu}_w\right)^2 \\
&= \left(\boldsymbol{\mu}_v^\top \boldsymbol{\mu}_w\right)^2 + \operatorname{tr}\left(\boldsymbol{\mu}_w\boldsymbol{\mu}_w^\top \boldsymbol{\Sigma}_v\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}_w\boldsymbol{\mu}_v\boldsymbol{\mu}_v^\top\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}_w\boldsymbol{\Sigma}_v\right) - \left(\boldsymbol{\mu}_v^\top \boldsymbol{\mu}_w\right)^2 \\
&= \boldsymbol{\mu}_v^\top \boldsymbol{\Sigma}_w \boldsymbol{\mu}_v + \boldsymbol{\mu}_w^\top \boldsymbol{\Sigma}_v \boldsymbol{\mu}_w + \operatorname{tr}\left(\boldsymbol{\Sigma}_w\boldsymbol{\Sigma}_v\right).
\end{aligned}$$

Using the Hadamard product identities $\boldsymbol{v}^\top(\boldsymbol{A} \odot \boldsymbol{B})\boldsymbol{w} = \operatorname{tr}\left(\boldsymbol{D}_v \boldsymbol{A} \boldsymbol{D}_w \boldsymbol{B}^\top\right)$ and $\boldsymbol{D}_v^\top \boldsymbol{A} \boldsymbol{D}_w = \boldsymbol{w}\boldsymbol{v}^\top \odot \boldsymbol{A}$, the covariance matrix identity $\boldsymbol{\Sigma}_w = \boldsymbol{D}_{\sigma_w} \boldsymbol{R}_w \boldsymbol{D}_{\sigma_w}$, and the interchangability of vectors in vector-diagonal matrix products $\boldsymbol{\mu}_v^\top \boldsymbol{D}_{\sigma_w} = \boldsymbol{\sigma}_w^\top \boldsymbol{D}_{\mu_v}$, the variance can be written as

$$\begin{aligned}
Var\left[\boldsymbol{v}^\top \boldsymbol{w}\right] &= \boldsymbol{\sigma}_w^\top \boldsymbol{D}_{\mu_v} \boldsymbol{R}_w \boldsymbol{D}_{\mu_v} \boldsymbol{\sigma}_w + \boldsymbol{\mu}_w^\top \boldsymbol{\Sigma}_v \boldsymbol{\mu}_w + \boldsymbol{\sigma}_w^\top (\boldsymbol{R}_w \odot \boldsymbol{\Sigma}_v) \boldsymbol{\sigma}_w \\
&= \boldsymbol{\sigma}_w^\top (\boldsymbol{R}_w \odot \boldsymbol{\mu}_v \boldsymbol{\mu}_v^\top) \boldsymbol{\sigma}_w + \boldsymbol{\sigma}_w^\top (\boldsymbol{R}_w \odot \boldsymbol{\Sigma}_v) \boldsymbol{\sigma}_w + \boldsymbol{\mu}_w^\top \boldsymbol{\Sigma}_v \boldsymbol{\mu}_w \\
&= \boldsymbol{\sigma}_w^\top (\boldsymbol{R}_w \odot \boldsymbol{\Omega}_v) \boldsymbol{\sigma}_w + \boldsymbol{\mu}_w^\top \boldsymbol{\Sigma}_v \boldsymbol{\mu}_w,
\end{aligned}$$

where $\boldsymbol{\Omega}_v = \boldsymbol{\mu}_v \boldsymbol{\mu}_v^\top + \boldsymbol{\Sigma}_v$.

A.2. Bias-variance trade-off solution

A.2.1. Linear-affine case

Substituting vectors of bias and volatility as functions of controls $\boldsymbol{c}$ given by Assumption 5 into the objective in (2) yields

$$\begin{aligned}
\Theta &= k_\varepsilon^2 \boldsymbol{c}^\top \boldsymbol{\Omega}_\zeta \boldsymbol{c} + k_\sigma^2 \boldsymbol{c}^\top \boldsymbol{D}_{\Omega_\zeta} \boldsymbol{c} + 2 k_{\sigma 0} k_\sigma \mathbf{1}^\top \boldsymbol{D}_{\Omega_\zeta} \boldsymbol{c} + k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{D}_{\Omega_\zeta} \mathbf{1} \\
&= k_\sigma^2 \boldsymbol{c}^\top (k_c^{-2} \boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}) \boldsymbol{c} + 2 k_{\sigma 0} k_\sigma \mathbf{1}^\top \boldsymbol{D}_{\Omega_\zeta} \boldsymbol{c} + k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{D}_{\Omega_\zeta} \mathbf{1}.
\end{aligned}$$

Hessian matrix is positive definite due to the full rank assumption Section **??** and given by

$$\frac{\partial \Theta}{\partial \boldsymbol{c}^2} = k_\varepsilon^2 \boldsymbol{\Omega}_\zeta + k_\sigma^2 \boldsymbol{D}_{\Omega_\zeta}.$$

Optimal controls from first order condition are

$$\begin{aligned}
\mathbf{0} = \frac{\partial \Theta}{\partial \boldsymbol{c}} &= 2k_\sigma^2 \boldsymbol{c}^\top (k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}) + 2k_{\sigma 0}k_\sigma \mathbf{1}^\top \boldsymbol{D}_{\Omega_\zeta} \\
\iff \boldsymbol{c}^* &= -k_\sigma^{-1} k_{\sigma 0} \left(k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}\right)^{-1} \boldsymbol{D}_{\Omega_\zeta}\mathbf{1},
\end{aligned}$$

and substituting back into the objective it simplifies to

$$\begin{aligned}
\chi = \Theta|_{\boldsymbol{c}=\boldsymbol{c}^*} &= k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{D}_{\Omega_\zeta} \left(k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}\right)^{-1} (k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}) \left(k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}\right)^{-1} \boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \\
&\quad - 2k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{D}_{\Omega_\zeta} \left(k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}\right)^{-1} \boldsymbol{D}_{\Omega_\zeta}\mathbf{1} + k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \\
&= -k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{D}_{\Omega_\zeta} \left(k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}\right)^{-1} \boldsymbol{D}_{\Omega_\zeta}\mathbf{1} + k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \\
&= -k_{\sigma 0}^2 \mathbf{1}^\top \left\{\boldsymbol{D}_{\Omega_\zeta} - \left(k_c^2 \boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1}\right)^{-1}\right\} \mathbf{1} + k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \\
&= k_{\sigma 0}^2 \mathbf{1}^\top \left\{k_c^2 \boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1}\right\}^{-1} \mathbf{1}.
\end{aligned}$$

Letting $\boldsymbol{X} := k_c^2 \boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1}$, optimal controls and cost of complexity are

$$\boldsymbol{c}^* = -k_\sigma^{-1} k_{\sigma 0} \left\{\boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1} \boldsymbol{X}^{-1}\right\} \mathbf{1}, \quad \chi = \Theta|_{\boldsymbol{c}=\boldsymbol{c}^*} = k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{X}^{-1} \mathbf{1},$$

and bias and volatility

$$\begin{aligned}
\boldsymbol{\varepsilon}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} &= k_\varepsilon \left(-k_\sigma^{-1} k_{\sigma 0} \left\{\boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1} \boldsymbol{X}^{-1}\right\} \mathbf{1}\right) = -k_c^{-1} k_{\sigma 0} \left\{\boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1} \boldsymbol{X}^{-1}\right\} \mathbf{1}, \\
\boldsymbol{\sigma}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} &= k_\sigma^{-1} \left(-k_\sigma^{-1} k_{\sigma 0} \left\{\boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1} \boldsymbol{X}^{-1}\right\} \mathbf{1}\right) + k_{\sigma 0}\mathbf{1} = k_{\sigma 0} \boldsymbol{D}_{\Omega_\zeta}^{-1} \boldsymbol{X}^{-1} \mathbf{1}.
\end{aligned}$$

A.2.2. Feasibility condition simplification and positive optimal bias

Intermediary steps for the condition in Assumption 6. Notice that $k_{\sigma 0} > 0$ and by the Woodbury matrix identity

$$\boldsymbol{X}^{-1} = \boldsymbol{D}_{\Omega_\zeta} - \boldsymbol{D}_{\Omega_\zeta}(k^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta})^{-1}\boldsymbol{D}_{\Omega_\zeta} \implies \boldsymbol{D}_{\Omega_\zeta}^{-1}\boldsymbol{X}^{-1} = \boldsymbol{I} - (k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta})^{-1}\boldsymbol{D}_{\Omega_\zeta}.$$

The non-zero condition on the vector of volatilities of the estimator $\hat{\boldsymbol{\beta}}$ at the optimum $\boldsymbol{c}^*$ can be rewritten as

$$\boldsymbol{\sigma}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} = k_{\sigma 0}\boldsymbol{D}_\Omega^{-1}\boldsymbol{X}^{-1}\mathbf{1} \geq \mathbf{0} \iff \left\{\boldsymbol{I} - (k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta})^{-1}\boldsymbol{D}_{\Omega_\zeta}\right\}\mathbf{1} \geq \mathbf{0}$$
$$\iff (k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta})\mathbf{1} \geq \boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \iff \boldsymbol{\Omega}_\zeta\mathbf{1} \geq \mathbf{0}.$$

For the bias, notice $-k_c^{-1}k_{\sigma 0} > 0$ so

$$\boldsymbol{\varepsilon}_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} = \left\{\boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1}\boldsymbol{X}^{-1}\right\}\mathbf{1} \geq \mathbf{0} \iff \mathbf{1} \geq \boldsymbol{D}_{\Omega_\zeta}^{-1}\boldsymbol{X}^{-1}\mathbf{1} \iff \boldsymbol{X}\boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \geq \mathbf{1}$$
$$\iff (k_c^2\boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1})\boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \geq \mathbf{1} \iff k_c^2\boldsymbol{\Omega}_\zeta^{-1}\boldsymbol{D}_{\Omega_\zeta}\mathbf{1} + \mathbf{1} \geq \mathbf{1}$$
$$\iff \boldsymbol{D}_{\Omega_\zeta}\mathbf{1} + k_c^{-2}\boldsymbol{\Omega}_\zeta\mathbf{1} \geq k_c^{-2}\boldsymbol{\Omega}_\zeta\mathbf{1} \iff \boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \geq \mathbf{0},$$

and the bias is always positive.

A.2.3. Affine-affine case

Relax Assumption 5 to allow bias to have intercept $k_{\varepsilon 0}$, let $k_{c0} := k_{\sigma 0}/k_{\varepsilon 0}$, $\tilde{\boldsymbol{\Omega}}_\zeta := \boldsymbol{D}_{\Omega_\zeta}^{-\frac{1}{2}}\boldsymbol{\Omega}_\zeta\boldsymbol{D}_{\Omega_\zeta}^{-\frac{1}{2}}$, $\tilde{\boldsymbol{\Omega}}_{\zeta*1} := \tilde{\boldsymbol{\Omega}}_\zeta + k_c k_{c0}\boldsymbol{I}$, $\tilde{\boldsymbol{\Omega}}_{\zeta*0} := \tilde{\boldsymbol{\Omega}}_\zeta + k_{c0}^2\boldsymbol{I}$, then optimal control and objective at solution are given by

$$\boldsymbol{c}^* = -\frac{k_{\varepsilon 0}}{k_\varepsilon}\boldsymbol{D}_\Omega^{-\frac{1}{2}}\tilde{\boldsymbol{\Omega}}_{\zeta*}^{-1}\tilde{\boldsymbol{\Omega}}_{\zeta*1}\boldsymbol{D}_\Omega^{\frac{1}{2}}\mathbf{1}, \quad \chi = k_{\varepsilon 0}^2\mathbf{1}^\top\boldsymbol{D}_\Omega^{\frac{1}{2}}\left\{\tilde{\boldsymbol{\Omega}}_{\zeta*0} - \tilde{\boldsymbol{\Omega}}_{\zeta*1}\tilde{\boldsymbol{\Omega}}_{\zeta*}^{-1}\tilde{\boldsymbol{\Omega}}_{\zeta*1}\right\}\boldsymbol{D}_\Omega^{\frac{1}{2}}\mathbf{1},$$

based on derivations analogous to Appendix A.2.1. Notice that substituting optimal control into the objective the expression for χ follows from

$$\begin{aligned}\chi = \mathbf{1}^\top\Bigg\{& k_\varepsilon^2\frac{k_{\varepsilon 0}^2}{k_\varepsilon^2}\boldsymbol{D}_\Omega^{\frac{1}{2}}\tilde{\boldsymbol{\Omega}}_{\zeta*0}\tilde{\boldsymbol{\Omega}}_{\zeta*}^{-1}\tilde{\boldsymbol{\Omega}}_\zeta\tilde{\boldsymbol{\Omega}}_{\zeta*}^{-1}\tilde{\boldsymbol{\Omega}}_{\zeta*0}\boldsymbol{D}_\Omega^{\frac{1}{2}} + k_\sigma^2\frac{k_{\varepsilon 0}^2}{k_\varepsilon^2}\boldsymbol{D}_\Omega^{\frac{1}{2}}\tilde{\boldsymbol{\Omega}}_{\zeta*0}\tilde{\boldsymbol{\Omega}}_{u*}^{-1}\boldsymbol{I}\tilde{\boldsymbol{\Omega}}_{\zeta*}^{-1}\tilde{\boldsymbol{\Omega}}_{\zeta*0}\boldsymbol{D}_\Omega^{\frac{1}{2}}\\
&- 2k_\varepsilon k_{\varepsilon 0}\frac{k_{\varepsilon 0}}{k_\varepsilon}\boldsymbol{D}_\Omega^{\frac{1}{2}}\tilde{\boldsymbol{\Omega}}_\zeta\tilde{\boldsymbol{\Omega}}_{\zeta*}^{-1}\tilde{\boldsymbol{\Omega}}_{u*0}\boldsymbol{D}_\Omega^{\frac{1}{2}} - k_\sigma k_{\sigma 0}\frac{k_{\varepsilon 0}}{k_\varepsilon}\boldsymbol{D}_\Omega^{\frac{1}{2}}\boldsymbol{I}\tilde{\boldsymbol{\Omega}}_{\zeta*}^{-1}\tilde{\boldsymbol{\Omega}}_{\zeta*0}\boldsymbol{D}_\Omega^{\frac{1}{2}}\\
&+ k_{\varepsilon 0}^2\boldsymbol{D}_\Omega^{\frac{1}{2}}\tilde{\boldsymbol{\Omega}}_\zeta\boldsymbol{D}_\Omega^{\frac{1}{2}} + \frac{k_{\varepsilon 0}^2}{k_{\varepsilon 0}^2}k_{\sigma 0}^2\boldsymbol{D}_\Omega\Bigg\}\mathbf{1},\end{aligned}$$

by collecting terms in the first and second line.

A.3. Recursive formulation of $\boldsymbol{\Gamma}_{n_s}$ single signal

For every entry $\boldsymbol{\Gamma}_{i,j,n_s} = \gamma_{ij,n_s} = \boldsymbol{\rho}_{q_i s,n_s}\boldsymbol{R}_{s,n_s}^{-1}\boldsymbol{\rho}_{q_j s,n_s}^\top$, denoting the vector of correlations between s_{n_s} and other informed signals by $\boldsymbol{\rho}_{s,n_s} = Corr(\boldsymbol{s}_{n_s-1}, s_{n_s})$, and the correlation correction $\rho_{s,n_s|n_s-1} = 1 - \boldsymbol{\rho}_{s,n_s}^\top \boldsymbol{R}_{s,n_s-1}^{-1} \boldsymbol{\rho}_{s,n_s}$, block matrix inversion provides the decomposition

$$\begin{aligned}
\boldsymbol{R}_{s,n_s} &= \begin{pmatrix} \boldsymbol{R}_{s,n_s-1} & \boldsymbol{\rho}_{s,n_s} \\ \boldsymbol{\rho}_{s,n_s}^\top & 1 \end{pmatrix} \\
\implies \boldsymbol{R}_{s,n_s}^{-1} &= \begin{pmatrix} \boldsymbol{R}_{s,n_s-1}^{-1} & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{pmatrix} + \frac{1}{\rho_{s,n_s|n_s-1}} \begin{pmatrix} \boldsymbol{R}_{s,n-1}^{-1} \boldsymbol{\rho}_{s,n_s} \boldsymbol{\rho}_{s,n_s}^\top \boldsymbol{R}_{s,n-1}^{-1} & -\boldsymbol{R}_{s,n_s-1}^{-1} \boldsymbol{\rho}_{s,n_s} \\ -\boldsymbol{\rho}_{s,n_s}^\top \boldsymbol{R}_{s,n_s-1}^{-1} & 1 \end{pmatrix}.
\end{aligned}$$

The vector of correlations between a factor and all signals can be split into $\boldsymbol{\rho}_{q_i s,n_s}^\top = \begin{pmatrix} \boldsymbol{\rho}_{q_i s,n_s-1} & \rho_{q_i s_{n_s}} \end{pmatrix}$, let $\phi_{i,n_s} := \boldsymbol{\rho}_{q_i s,n-1}^\top \boldsymbol{R}_{s,n_s-1}^{-1} \boldsymbol{\rho}_{s,n_s} - \rho_{q_i s_{n_s}}$ and the quadratic form γ_{ij,n_s} can be rewritten as

$$\begin{aligned}
\gamma_{ij,n} = \boldsymbol{\rho}_{q_i s,n_s-1}^\top \boldsymbol{R}_{s,n_s-1}^{-1} \boldsymbol{\rho}_{q_j s,n_s-1} + \frac{1}{\rho_{s,n_s|n_s-1}} \Big\{ & \boldsymbol{\rho}_{q_i s,n-1}^\top \boldsymbol{R}_{s,n-1}^{-1} \boldsymbol{\rho}_{s,n} \boldsymbol{\rho}_{s,n}^\top \boldsymbol{R}_{s,n_s-1}^{-1} \boldsymbol{\rho}_{q_j s,n_s-1} + \rho_{q_i s_n} \rho_{q_j s_n} \\
& - \rho_{q_j s_n} \boldsymbol{\rho}_{q_i s,n-1}^\top \boldsymbol{R}_{s,n-1}^{-1} \boldsymbol{\rho}_{s,n} - \rho_{q_i s_n} \boldsymbol{\rho}_{s,n}^\top \boldsymbol{R}_{s,n-1}^{-1} \boldsymbol{\rho}_{q_j s,n-1} \Big\} = \gamma_{ij,n_s-1} + \frac{\phi_{i,n_s} \phi_{j,n_s}}{\rho_{s,n_s|n_s-1}}.
\end{aligned}$$

Since the correlation correction $\rho_{s,n_s|n_s-1}$ is the same across entries, defining the vector $\boldsymbol{\phi}_n^\top = \begin{pmatrix} \phi_{1,n_s} & \phi_{2,n_s} & \dots & \phi_{n_q,n_s} \end{pmatrix}$, the full matrix can be written recursively as

$$\boldsymbol{\Gamma}_{n_s} = \boldsymbol{\Gamma}_{n_s-1} + \frac{1}{\rho_{s,n_s|n_s-1}} \boldsymbol{\phi}_{n_s} \boldsymbol{\phi}_{n_s}^\top$$

The diagonal elements of $\boldsymbol{\Gamma}_{n_s}$ is weakly increasing in n_s as

$$\gamma_{ii,n_s} = \gamma_{ii,n_s-1} + \frac{\phi_{i,n_s}^2}{\rho_{s,n_s|n_s-1}} \geq \gamma_{ii,n_s-1},$$

and the sum over all entries is as well, since the outer product $\boldsymbol{\phi}_{n_s} \boldsymbol{\phi}_{n_s}^\top$ is positive semi-definite, formally

$$\mathbf{1}^\top (\boldsymbol{\Gamma}_{n_s} - \boldsymbol{\Gamma}_{n_s-1}) \mathbf{1} = \frac{1}{\rho_{s,n_s|n_s-1}} \mathbf{1}^\top \boldsymbol{\phi}_{n_s} \boldsymbol{\phi}_{n_s}^\top \mathbf{1} \geq 0.$$

A.4. Cost and benefit of complexity single signal

The impact of adding another signal is always to weakly increase χ as can be demonstrated by the positive semi-definiteness of the difference $\boldsymbol{X}_n^{-1} - \boldsymbol{X}_{n-1}^{-1}$. By properties of symmetric positive definite matrices,[17] the difference is semi-positive definite if the differ-

[17]For symmetric positive definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with the same dimensions it is the case that

$$\boldsymbol{v}^\top (\boldsymbol{A} - \boldsymbol{B}) \boldsymbol{v} \geq 0 \implies \boldsymbol{v}^\top (\boldsymbol{B}^{-1} - \boldsymbol{A}^{-1}) \boldsymbol{v} \geq 0 \ \forall \boldsymbol{v},$$

ence $\boldsymbol{X}_{n-1} - \boldsymbol{X}_n$ is. Let $\boldsymbol{D}_{\phi,n_s} = \rho^{-1}_{s,n_s|n_s-1} \operatorname{diag}\left(\boldsymbol{\phi}_{n_s}\boldsymbol{\phi}^\top_{n_s}\right)$, then explicit calculation using the Sherman-Morrison formula for the inverse of the sum of a positive definite matrix and the outer product of vectors of the difference yields

$$
\begin{aligned}
&\boldsymbol{X}_{n-1} - \boldsymbol{X}_n = \\
&\quad k_c^2 \boldsymbol{\Omega}^{-1}_{\zeta,n_s-1} + \boldsymbol{D}^{-1}_{\Omega_\zeta,n_s-1} - k_c^2 \boldsymbol{\Omega}^{-1}_{\zeta,n_s} - \boldsymbol{D}^{-1}_{\Omega_\zeta,n_s} = k_c^2\left(\boldsymbol{\Omega}^{-1}_{\zeta,n_s-1} - \boldsymbol{\Omega}^{-1}_{\zeta,n_s}\right) + \boldsymbol{D}^{-1}_{\Omega_\zeta,n_s-1} - \boldsymbol{D}^{-1}_{\Omega_\zeta,n_s} \\
&\qquad = k_c^2\left[\boldsymbol{\Omega}^{-1}_{\zeta,n_s-1} - \left(\boldsymbol{\Omega}^{-1}_{\zeta,n_s-1} - \frac{\rho^{-1}_{s,n_s|n_s-1}\boldsymbol{\Omega}^{-1}_{\zeta,n_s-1}\boldsymbol{D}_{\sigma_q}\boldsymbol{\phi}_{n_s}\boldsymbol{\phi}^\top_{n_s}\boldsymbol{D}_{\sigma_q}\boldsymbol{\Omega}^{-1}_{\zeta,n_s-1}}{1 + \rho^{-1}_{s,n|n-1}\left(\boldsymbol{\sigma}_q \odot \boldsymbol{\phi}_{n_s}\right)^\top \boldsymbol{\Omega}^{-1}_{\zeta,n_s-1}\left(\boldsymbol{\sigma}_q \odot \boldsymbol{\phi}_{n_s}\right)}\right)\right] \\
&\qquad\quad + \boldsymbol{D}^{-1}_{\Omega_\zeta,n_s-1} - \left[\boldsymbol{D}^{-1}_{\Omega_\zeta,n_s-1} - \left(\boldsymbol{D}_{\Omega_\zeta,n_s} + \boldsymbol{D}_{\Omega_\zeta,n_s}\boldsymbol{D}^{-1}_{\sigma_q}\boldsymbol{D}^{-1}_{\phi,n_s}\boldsymbol{D}^{-1}_{\sigma_q}\boldsymbol{D}_{\Omega,n}\right)^{-1}\right] \\
&\qquad = k_c^2 \frac{\rho^{-1}_{s,n_s|n_s-1}\boldsymbol{\Omega}^{-1}_{\zeta,n_s-1}\boldsymbol{D}_{\sigma_q}\boldsymbol{\phi}_{n_s}\boldsymbol{\phi}^\top_{n_s}\boldsymbol{D}_{\sigma_q}\boldsymbol{\Omega}^{-1}_{\zeta,n_s-1}}{1 + \rho^{-1}_{s,n_s|n_s-1}\left(\boldsymbol{\sigma}_q \odot \boldsymbol{\phi}_{n_s}\right)^\top \boldsymbol{\Omega}^{-1}_{\zeta,n_s-1}\left(\boldsymbol{\sigma}_q \odot \boldsymbol{\phi}_{n_s}\right)} \\
&\qquad\quad + \left(\boldsymbol{D}_{\Omega_\zeta,n_s} + \boldsymbol{D}_{\Omega_\zeta,n_s}\boldsymbol{D}^{-1}_{\sigma_q}\boldsymbol{D}^{-1}_{\phi,n_s}\boldsymbol{D}^{-1}_{\sigma_q}\boldsymbol{D}_{\Omega_\zeta,n_s}\right)^{-1},
\end{aligned}
$$

which, as a sum of (semi-)positive definite matrices, is semi-positive definite.

A.5. Recursive formulation of $\boldsymbol{\Gamma}_{n_s}$ multiple signals

For every entry $\boldsymbol{\Gamma}_{i,j,n_s} = \gamma_{ij,n_s} = \boldsymbol{\rho}_{q_i s,n_s}\boldsymbol{R}^{-1}_{s,n_s}\boldsymbol{\rho}^\top_{q_j s,n_s}$, denoting the matrix of correlations between $\boldsymbol{s}_{n_s+}$ and other informed signals by $\boldsymbol{R}_{n_s-n_s+} = Corr(\boldsymbol{s}_{I,n_s}, \boldsymbol{s}_{n_s+})$, and the correlation correction $\boldsymbol{R}_{n_s+|n_s-} = \boldsymbol{R}_{s,n_s+} - \boldsymbol{R}^\top_{n_s-n_s+}\boldsymbol{R}^{-1}_{s,n_s-}\boldsymbol{R}_{n_s-n_s+}$, block matrix inversion provides the decomposition

$$
\begin{aligned}
\boldsymbol{R}_{s,n_s} &= \begin{pmatrix} \boldsymbol{R}_{s,n_s-} & \boldsymbol{R}_{n_s-n_s+} \\ \boldsymbol{R}^\top_{n_s-n_s+} & \boldsymbol{R}_{s,n_s+} \end{pmatrix} \\
\implies \boldsymbol{R}^{-1}_{s,n_s} &= \begin{pmatrix} \boldsymbol{R}^{-1}_{s,n_s-} & \boldsymbol{0} \\ \boldsymbol{0}^\top & \boldsymbol{0}\boldsymbol{0}^\top \end{pmatrix} \\
&\quad + \begin{pmatrix} \boldsymbol{R}^{-1}_{s,n_s-}\boldsymbol{R}_{n_s-n_s+}\boldsymbol{R}^{-1}_{n_s+|n_s-}\boldsymbol{R}^\top_{n_s-n_s+}\boldsymbol{R}^{-1}_{s,n_s-} & -\boldsymbol{R}^{-1}_{s,n_s-}\boldsymbol{R}_{n_s-n_s+}\boldsymbol{R}^{-1}_{n_s+|n_s-} \\ -\boldsymbol{R}^{-1}_{n_s+|n_s-}\boldsymbol{R}^\top_{n_s-n_s+}\boldsymbol{R}^{-1}_{s,n_s-} & \boldsymbol{R}^{-1}_{n_s+|n_s-} \end{pmatrix}.
\end{aligned}
$$

The vector of correlations between a factor and all signals can be split into $\boldsymbol{\rho}^\top_{q_i s,n_s} = \begin{pmatrix}\boldsymbol{\rho}^\top_{q_i s,n_s-} & \boldsymbol{\rho}^\top_{q_i s,n_s+}\end{pmatrix}$, let $\boldsymbol{\phi}^\top_{i,n_s} := \boldsymbol{\rho}^\top_{q_i s,n_s-}\boldsymbol{R}^{-1}_{s,n_s-}\boldsymbol{R}_{n_s-n_s+} - \boldsymbol{\rho}^\top_{q_i s,n_s+}$ and the quadratic form

i.e. if the difference $\boldsymbol{A} - \boldsymbol{B}$ is positive semi-definite, the difference $\boldsymbol{B}^{-1} - \boldsymbol{A}^{-1}$ is positive semi-definite.

γ_{ij,n_s} can be rewritten as

$$\begin{aligned}
\gamma_{ij,n} &= \boldsymbol{\rho}_{q_i s,n_s-}^{\top} \boldsymbol{R}_{s,n_s-1}^{-1} \boldsymbol{\rho}_{q_j s,n_s-} \\
&\quad + \boldsymbol{\rho}_{q_i s,n_s-}^{\top} \boldsymbol{R}_{s,n_s-}^{-1} \boldsymbol{R}_{n_s-n_s+} \boldsymbol{R}_{n_s+|n_s-}^{-1} \boldsymbol{R}_{n_s-n_s+}^{\top} \boldsymbol{R}_{s,n_s-}^{-1} \boldsymbol{\rho}_{q_i s,n_s-} \\
&\quad - \boldsymbol{\rho}_{q_i s,n_s-}^{\top} \boldsymbol{R}_{s,n_s-}^{-1} \boldsymbol{R}_{n_s-n_s+} \boldsymbol{R}_{n_s+|n_s-}^{-1} \boldsymbol{\rho}_{q_i s,n_s+} \\
&\quad - \boldsymbol{\rho}_{q_i s,n_s+}^{\top} \boldsymbol{R}_{n_s+|n_s-}^{-1} \boldsymbol{R}_{n_s-n_s+}^{\top} \boldsymbol{R}_{s,n_s-}^{-1} \boldsymbol{\rho}_{q_i s,n_s-} \\
&\quad + \boldsymbol{\rho}_{q_i s,n_s+}^{\top} \boldsymbol{R}_{n_s+|n_s-}^{-1} \boldsymbol{\rho}_{q_i s,n_s+} \\
&= \gamma_{ij,n_s-1} + \boldsymbol{\phi}_{i,n_s}^{\top} \boldsymbol{R}_{s,n_s+|n_s-}^{-1} \boldsymbol{\phi}_{j,n_s}.
\end{aligned}$$

Since the correlation correction matrix $\boldsymbol{R}_{n_s+|n_s-}^{-1}$ is the same across entries, defining the matrix $\boldsymbol{\Phi}_{n_s} = \boldsymbol{R}_{qs,n_s-} \boldsymbol{R}_{s,n_s-}^{-1} \boldsymbol{R}_{n_s-n_s+} - \boldsymbol{R}_{qs,n_s+}$, where the rows are $\boldsymbol{\phi}_{i,n_s}^{\top}$, the full matrix can be written recursively as

$$\boldsymbol{\Gamma}_{n_s} = \boldsymbol{\Gamma}_{n_s-} + \boldsymbol{\Phi}_{n_s} \boldsymbol{R}_{s,n_s+|n_s-}^{-1} \boldsymbol{\Phi}_{n_s}^{\top}.$$

The diagonal elements of $\boldsymbol{\Gamma}_{n_s}$ are weakly increasing in n_s as $\boldsymbol{R}_{s,n_s+|n_s-}^{-1}$ is positive semi-definite so

$$\gamma_{ii,n_s} = \gamma_{ii,n_s-} + \boldsymbol{\phi}_{i,n_s} \boldsymbol{R}_{s,n_s+|n_s-}^{-1} \boldsymbol{\phi}_{i,n_s}^{\top} \geq \gamma_{ii,n_s-},$$

and the sum over all entries is as well, since $\boldsymbol{\Gamma}_{n_s} - \boldsymbol{\Gamma}_{n_s-} = \boldsymbol{\Phi}_{n_s} \boldsymbol{R}_{s,n_s+|n_s-}^{-1} \boldsymbol{\Phi}_{n_s}^{\top}$ is positive semi-definite.

A.6. Cost and benefit of complexity multiple signals

The impact of adding another group of signal is always to weakly increase χ as can be demonstrated by the positive semi-definiteness of the difference $\boldsymbol{X}_{n_s+}^{-1} - \boldsymbol{X}_{n_s-}^{-1}$. By properties of symmetric positive definite matrices,[18] the difference is semi-positive definite if the difference $\boldsymbol{X}_{n_s-} - \boldsymbol{X}_{n_s+}$ is. Explicit calculation using the Sherman-Morrison formula for the inverse of the sum of a positive definite matrix and the outer product of vectors of the

[18] For symmetric positive definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with the same dimensions it is the case that

$$\boldsymbol{v}^{\top}(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{v} \geq 0 \implies \boldsymbol{v}^{\top}(\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1})\boldsymbol{v} \geq 0 \; \forall \boldsymbol{v},$$

i.e. if the difference $\boldsymbol{A}-\boldsymbol{B}$ is positive semi-definite, the difference $\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1}$ is positive semi-definite.

difference yields

$$
\begin{aligned}
&\boldsymbol{X}_{n_s-} - \boldsymbol{X}_{n_s+} = \\
&\quad k_c^2 \boldsymbol{\Omega}_{\zeta,n_s-}^{-1} + \boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1} - k_c^2 \boldsymbol{\Omega}_{\zeta,n_s}^{-1} - \boldsymbol{D}_{\Omega_\zeta,n_s}^{-1} = k_c^2 \left(\boldsymbol{\Omega}_{\zeta,n_s-}^{-1} - \boldsymbol{\Omega}_{\zeta,n_s}^{-1} \right) + \boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1} - \boldsymbol{D}_{\Omega_\zeta,n_s}^{-1} \\
&\qquad = k_c^2 \left[\boldsymbol{\Omega}_{\zeta,n_s-}^{-1} - \left(\boldsymbol{\Omega}_{\zeta,n_s-}^{-1} - \boldsymbol{\Omega}_{\zeta,n_s-}^{-1} \boldsymbol{\Phi}_{n_s}^\top \left(\boldsymbol{R}_{s,n_s+|n_s-} + \boldsymbol{\Phi}_{n_s}^\top \boldsymbol{\Omega}_{\zeta,n_s-}^{-1} \boldsymbol{\Phi}_{n_s} \right) \boldsymbol{\Phi}_{n_s}^\top \boldsymbol{\Omega}_{\zeta,n_s-}^{-1} \right) \right] \\
&\qquad + \boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1} - \left[\boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1} - \left(\boldsymbol{D}_{\Omega_\zeta,n_s} + \boldsymbol{D}_{\Omega_\zeta,n_s} \boldsymbol{D}_{\sigma_q}^{-1} \boldsymbol{D}_{\Phi,n_s}^{-1} \boldsymbol{D}_{\sigma_q}^{-1} \boldsymbol{D}_{\Omega,n} \right)^{-1} \right] \\
&\qquad = k_c^2 \boldsymbol{\Omega}_{\zeta,n_s-}^{-1} \boldsymbol{\Phi}_{n_s}^\top \left(\boldsymbol{R}_{s,n_s+|n_s-} + \boldsymbol{\Phi}_{n_s}^\top \boldsymbol{\Omega}_{\zeta,n_s-}^{-1} \boldsymbol{\Phi}_{n_s} \right) \boldsymbol{\Phi}_{n_s}^\top \boldsymbol{\Omega}_{\zeta,n_s-}^{-1} \\
&\qquad + \left(\boldsymbol{D}_{\Omega_\zeta,n_s} + \boldsymbol{D}_{\Omega_\zeta,n_s} \boldsymbol{D}_{\sigma_q}^{-1} \boldsymbol{D}_{\Phi,n_s}^{-1} \boldsymbol{D}_{\sigma_q}^{-1} \boldsymbol{D}_{\Omega_\zeta,n_s} \right)^{-1},
\end{aligned}
$$

which, as a sum of (semi-)positive definite matrices, is semi-positive definite.

A.7. Impact on including signals of improving technology

The matrix derivative

$$
\begin{aligned}
\frac{\partial \boldsymbol{X}_{n_s+}^{-1} - \boldsymbol{X}_{n_s-}^{-1}}{\partial k_c^2} &= -\boldsymbol{X}_{n_s+}^{-1} \frac{\partial \boldsymbol{X}_{n_s+}}{\partial k_c^2} \boldsymbol{X}_{n_s+}^{-1} + \boldsymbol{X}_{n_s-}^{-1} \frac{\boldsymbol{X}_{n_s-}}{\partial k_c^2} \boldsymbol{X}_{n_s-}^{-1} \\
&= -\boldsymbol{X}_{n_s+}^{-1} \boldsymbol{\Omega}_{\zeta,n_s+}^{-1} \boldsymbol{X}_{n_s+} + \boldsymbol{X}_{n_s-}^{-1} \boldsymbol{\Omega}_{\zeta,n_s-}^{-1} \boldsymbol{X}_{n_s-}^{-1}
\end{aligned}
$$

is positive semi-definite if the matrix difference

$$
\begin{aligned}
&\boldsymbol{X}_{n_s+} \boldsymbol{\Omega}_{\zeta,n_s+} \boldsymbol{X}_{n_s+} - \boldsymbol{X}_{n_s-} \boldsymbol{\Omega}_{\zeta,n_s-} \boldsymbol{X}_{n_s-} \\
&\quad = k_c^4 \left\{ \boldsymbol{\Omega}_{\zeta,n_s+}^{-1} - \boldsymbol{\Omega}_{\zeta,n_s-}^{-1} \right\} + \left\{ \boldsymbol{D}_{\Omega_\zeta,n_s+}^{-1} \boldsymbol{\Omega}_{\zeta,n_s+} \boldsymbol{D}_{\Omega_\zeta,n_s+}^{-1} - \boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1} \boldsymbol{\Omega}_{\zeta,n_s-} \boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1} \right\} \\
&\quad + k_c^2 \left\{ \boldsymbol{\Omega}_{\zeta,n_s+} \boldsymbol{D}_{\Omega_\zeta,n_s+}^{-1} + \boldsymbol{D}_{\Omega_\zeta,n_s+}^{-1} \boldsymbol{\Omega}_{\zeta,n_s+} - \boldsymbol{\Omega}_{\zeta,n_s-} \boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1} - \boldsymbol{D}_{\Omega_\zeta,n_s-}^{-1} \boldsymbol{\Omega}_{\zeta,n_s-} \right\}
\end{aligned}
$$

is positive definite. This is always eventually true as $k_c^2 \to \infty$ and only the first term remains since $\boldsymbol{\Omega}_{\zeta,n_s-} - \boldsymbol{\Omega}_{\zeta,n_s+}$ is positive definite.

A.8. Objective measure is the limiting measure

This is true as the cost of complexity is reduced to zero as estimation technology parameter goes to infinity $k_c^2 \to \infty$ (k_c is the ratio of slope parameters in the bias variance trade-off minimization, see Section 2.3). This can seen by taking the limit of a slightly rearranged cost of complexity χ

$$\lim_{k_c^2 \to} \chi = \lim_{k_c^2 \to} k_c^{-2} k_{\sigma 0}^2 \mathbf{1}^\top \left\{ \boldsymbol{\Omega}_\zeta^{-1} + k_c^{-2} \boldsymbol{D}_{\Omega_\zeta}^{-1} \right\}^{-1} \mathbf{1} = 0 \times k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{\Omega}_\zeta \mathbf{1} = 0.$$

Appendix B. Heterogeneous agents

B.1. Profit function from budget constraint

Normalize the price and pay-off of the risk free asset to one (equivalent to a risk free rate of zero). The value of the position in the risk free asset is its size, denote it by B_i. With the initial value of position in the risky asset given by position times price $\delta_i p$, the budget constraint that the total value of investments cannot exceed initial wealth w_{0i} is $w_{0i} \geq B_i + \delta_i p$. For a utility function increasing in wealth the budget constraint binds, and it follows that $B_i = w_{0i} - \delta_i p$. Therefore, after-trade wealth is

$$w_{1i} = B_i + \delta_i y = w_{0i} - \delta_i p + \delta_i y = w_{0i} + \delta_i (y - p).$$

It can be shown that in determining position δ_i, for an investor with CARA utility of after-trade wealth, initial wealth (endowment) can be normalized to one without loss of generality, see Breon-Drish (2015). This holds as well for an investor maximizing a mean-variance criterion as initial wealth drops out of the first order condition because it enters wealth additively, which extends to the mean-mean squared error criterion described in the main text. In all three cases optimizing the profit function $\pi_i = \delta_i(y-p)$ is equivalent to optimizing after-trade wealth.

B.2. Predicting prediction

To see that $E[y|s_U, \boldsymbol{\beta}] = E[\hat{y}_I|s_U, \boldsymbol{\beta}]$, notice that for known $\boldsymbol{\beta}$

$$\begin{aligned} Cov[y, s_U|\boldsymbol{\beta}] &= Cov[y, \hat{y}_I|\boldsymbol{\beta}] = \boldsymbol{\beta}^\top Cov[\boldsymbol{q}, \boldsymbol{u}]\boldsymbol{\beta} = \boldsymbol{\beta}^\top Cov[\boldsymbol{q}, \boldsymbol{s}_I]\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{qs}^\top \boldsymbol{\beta} = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{qs}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{qs}^\top \boldsymbol{\beta} \\ &= \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{qs}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_s\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{qs}^\top \boldsymbol{\beta} = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{qs}\boldsymbol{\Sigma}_s^{-1} Var[\boldsymbol{s}_I]\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{qs}^\top \boldsymbol{\beta} = Var[\hat{y}_I|\boldsymbol{\beta}], \end{aligned}$$

and

$$E[y|\boldsymbol{\beta}] = \boldsymbol{\beta}^\top \boldsymbol{\mu}_q = \boldsymbol{\beta}^\top \{\boldsymbol{\mu}_q + \boldsymbol{\Sigma}_{qs}\boldsymbol{\Sigma}_s^{-1}(E[\boldsymbol{s}_I] - \boldsymbol{\mu}_s)\} = E[\hat{y}_I|\boldsymbol{\beta}] = E[s_U|\boldsymbol{\beta}],$$

so

$$E[y|s_U,\boldsymbol{\beta}] = E[\hat{y}_I|\boldsymbol{\beta}] + \frac{Var[\hat{y}_I|\boldsymbol{\beta}]}{Var[s_U|\boldsymbol{\beta}]}(s_U - E[s_U|\boldsymbol{\beta}]) = E[\hat{y}_I|s_U,\boldsymbol{\beta}].$$

B.3. Demand: robust profit maximization objective

I assume that investors optimize an extended mean-variance objective that consist of the expectation of the scaled profit function $\tilde{\pi}_i(y) := \delta_i\alpha_i(y-p)$ applied to the prediction $\hat{y}_i$ and an uncertainty-adjustment for the fact that investors optimize estimated rather than true profits extended from unconditional variance to unconditional mean squared error. In this two period model, optimizing over profits corresponds to optimizing over second period wealth (see Appendix B.1). The specification reflects the fact that investors ultimately care about true profits but are averse to the risk in the pay-off as well as the model uncertainty. To account for both sources of randomness, the uncertainty-adjustment is based on the unconditional mean squared error, and, for consistency with the problem faced by investors with CARA-utility facing a Gaussian gamble, it is half, i.e. $\frac{1}{2}\delta_i^2 E\left[(\tilde{\pi}_i(y)-\tilde{\pi}_i(\hat{y}))^2\right] = \frac{1}{2}\alpha_i^2 E\left[(y-\hat{y}_i)^2\right]$. Formally, demand from optimizing the objective function yields

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

where the expression follows from a reorganization of the first-order condition analogous to a classic mean-variance optimization and the second-order condition is satisfied due to the positivity of the mean squared error. If investors know the true model they act as mean-variance optimizers since the covariance between the pay-off and the predictor is the variance of the predictor $Cov[y,\hat{y}_i] = Var[\hat{y}_i]$ and the predictor is unbiased, $E[y] - E[\hat{y}_i]$, so the mean squared of the predictor equals the conditional variance of the pay-off given the predictor, i.e.

$$E\left[(y-\hat{y}_i)^2\right] = Var[y] - Var[\hat{y}_i] = Var[y|\hat{y}_i].$$

Without the noisy estimation introduced in Section 2.1, my specification of demand is, as was the case for uninformed inference in Section 3.1, simply a re-formulation of the baseline model provided by Grossman and Stiglitz (1980), the classic mean-variance criterion for utility optimization.

B.4. Demand: CARA-utility with ambiguity aversion

For the informed investors I in Section 3.1, their linear demand function is the demand of investors with CARA-utility and risk tolerance α_I, who performs a maximization of utility of final wealth w_{1I} over demand δ_I given price p that is robust to miscalculated risk. Conditional on the estimate $\hat{\boldsymbol{\beta}}$, the pay-off y is normally distributed, which extends to final wealth w_{I1}. Utility of final wealth is log-normal and expected utility is

$$E[U_I] = E\left[-e^{-\alpha_I^{-1} w_{1I}} \middle| \hat{\boldsymbol{\beta}}, \boldsymbol{s}, p\right] = -e^{-\alpha_I^{-1} E\left[w_{1I} \middle| \hat{\boldsymbol{\beta}}, \boldsymbol{s}, p\right] + \frac{1}{2}\alpha_I^{-2} Var\left[w_{1I} \middle| \hat{\boldsymbol{\beta}}, \boldsymbol{s}, p\right]}.$$

Maximizing the negative exponential is equivalent to minimizing its exponent, which again can be turned into a maximization by swapping the sign on the objective. Starting wealth can be normalized to zero without loss of generality, so final wealth is $w_{1I} = \delta_I(y - p)$, see Appendix B.1. As in the main text, the conditional expectation of the pay-off is the predictor $\hat{y}_I$. Denote the conditional variance of the pay-off by $\sigma^2_{\hat{y}I} \in \left[\underline{\sigma}^2_{\hat{y}I}, \overline{\sigma}^2_{\hat{y}I}\right]$, where the interval is the set of multiple priors of a maxmin expected utility model in the tradition of Gilboa and Schmeidler (1989).

$$\max_{\delta_I} \min_{\sigma^2_{\hat{y}I}} \alpha_I^{-1} \delta_I(\hat{y}_I - p) - \delta_I^2 \frac{1}{2} \alpha_I^{-2} \sigma^2_{\hat{y}I}, \quad s.t. \quad \sigma^2_{\hat{y}I} \in \left[\underline{\sigma}^2_{\hat{y}I}, \overline{\sigma}^2_{\hat{y}I}\right].$$

The first order condition of the minimization is the expression $-\delta_i^2 \alpha_i^{-2}/2$, which is always negative, meaning that the unconstrained solution would be positive infinity and the constrained solution is the upper bound. Substituting the upper bound into the objective, differentiating, and solving for the optimal position in the risky asset yields a result with a familiar form $\delta_I^* = \alpha_I \overline{\sigma}^{-2}_{\hat{y}I}(\hat{y}_I - p)$. However, rather than being scaled by the inverse conditional variance, the position is scaled by the inverse worst-outcome variance. The product of ambiguity and ambiguity aversion, effective ambiguity, is pinned down by defining the upper bound of the multiple priors set. Setting it equal to the unconditional mean squared error, i.e. $\overline{\sigma}^2_{\hat{y}I} = E[(y - \hat{y}_I)^2]$, yields the demand function in the main text.

The assumption that the uninformed predictor is the best linear approximation of the the informed prediction $\hat{y}_I$ can be tightened to the uninformed investors approximating the distribution of $\hat{y}_I$ by a normal distribution. This stronger assumption also makes $\hat{y}_U$ the projection presented in the main text. Additionally, following the steps outlined for the informed investors, uninformed investors' demand function in Section 3.1 corresponds to CARA-utility with risk tolerance α_U optimizing over profit $\delta_U(\hat{y}_I - p)$ robust to miscalculation of risk. Formally, expected utility is

$$E[U_U] = E\left[-e^{-\alpha_U^{-1}\delta_U(\hat{y}_I - p)}\Big|s_U,p\right] = -e^{-\alpha_U^{-1}\delta_U E\left[\hat{y}_I\Big|\boldsymbol{s}_U,p\right]+\frac{1}{2}\alpha_U^{-2}Var\left[\hat{y}_I\Big|\boldsymbol{s}_U,p\right]}.$$

With predictor $\hat{y}_U = E\left[\hat{y}_I\big|\boldsymbol{s}_U, p\right]$ and conditional variance $Var\left[\hat{y}_I\big|\boldsymbol{s}_U, p\right] = \sigma^2_{\hat{y}U} \in \left[\underline{\sigma}^2_{\hat{y}U}, \overline{\sigma}^2_{\hat{y}U}\right]$ the robust optimization is

$$\max_{\delta_U}\min_{\sigma^2_{\hat{y}U}} \alpha_U^{-1}\delta_U(\hat{y}_U - p) - \delta_U^2\frac{1}{2}\alpha_U^{-2}\sigma^2_{\hat{y}U}, \quad s.t. \quad \sigma^2_{\hat{y}U} \in \left[\underline{\sigma}^2_{\hat{y}U}, \overline{\sigma}^2_{\hat{y}U}\right].$$

Setting the upper limit equal to the unconditional mean squared error, i.e. $\overline{\sigma}^2_{\hat{y}U} = E[(y - \hat{y}_U)^2]$, yield the demand function in the main text.

B.5. Uninformed mean squared error

The mean squared error of the predictor of the uninformed

$$\begin{aligned}
E[(y-\hat{y}_U)^2] &= Var[y] + Var[\hat{y}_U] - 2Cov[y,\hat{y}_U] + (E[y]-E[\hat{y}_U])^2 \\
&= Var[y] + \lambda_U^2\left\{Var[\hat{y}_I] + \psi_I^{-2}\sigma_z^2\right\} - 2\lambda_U Cov[y,\hat{y}_I] \\
&\quad + [(\boldsymbol{\beta} - \{(1-\lambda_U)\boldsymbol{\mu}_\beta + \lambda_U\boldsymbol{\mu}_\beta\})^\top\boldsymbol{\mu}_q]^2 \\
&= Var[y] + \lambda_U(Var[\hat{y}_I] - 2Cov[y,\hat{y}_I]) + (\boldsymbol{\varepsilon}_\beta^\top\boldsymbol{\mu}_q)^2 \\
&= Var[y] + \lambda_U(\chi - \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_\zeta\boldsymbol{\beta}) + (\boldsymbol{\varepsilon}_\beta^\top\boldsymbol{\mu}_q)^2 \\
&= (1-\lambda_U)\left\{Var[y] + (\boldsymbol{\varepsilon}_\beta^\top\boldsymbol{\mu}_q)^2\right\} + \lambda_U E[(y-\hat{y}_I)^2]
\end{aligned}$$

so a necessary and sufficient condition for higher lower squared error of informed vs uninformed is

$$\begin{aligned}
E[(y-\hat{y}_U)^2] > E[(y-\hat{y}_I)^2] &\iff \left\{Var[y] + (\boldsymbol{\varepsilon}_\beta^\top\boldsymbol{\mu}_q)^2\right\} > E[(y-\hat{y}_I)^2] \\
&\iff \chi < \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_\zeta\boldsymbol{\beta} + (\boldsymbol{\varepsilon}_\beta^\top\boldsymbol{\mu}_q)^2 = Var[E[y|\boldsymbol{\beta},\boldsymbol{s}_I]] + (\boldsymbol{\varepsilon}_\beta^\top\boldsymbol{\mu}_q)^2.
\end{aligned}$$

B.6. Variable share of informed investors

Market clearing $\ell_I\delta_I(p) + (1-\ell_I)\delta_U(p) = z$ so

$$s_U = p - \psi_I^{-1}\ell_I^{-1}(1-\ell_I)\delta_U(p) = \hat{y}_I - \psi_I^{-1}\ell_I^{-1}z$$

and

$$p = \frac{\ell_I \psi_I \hat{y}_I + (1-\ell_I)\psi_U \hat{y}_U - z}{\ell_I \psi_I + (1-\ell_I)\psi_U} = \frac{\{\ell_I \psi_I + (1-\ell_I)\lambda_U \psi_U\} s_U + (1-\ell_I)(1-\lambda_U)\psi_U E[y|\boldsymbol{c}^*]}{\ell_I \psi_I + (1-\ell_I)\psi_U}$$
$$= (1-\lambda_p)E[y|\boldsymbol{c}^*] + \lambda_p s_U \quad s.t. \quad \lambda_p = \frac{\ell_I \psi_I + (1-\ell_I)\lambda_U \psi_U}{\ell_I \psi_I + (1-\ell_I)\psi_U}$$

so the changes to equilibrium outcomes are captured by price responsiveness λ_p, uninformed responsiveness λ_U, and uninformed signal s_U.

Appendix C. Predictions

C.1. Short-term price reversals

Price and stochastic supply. Moments: variance

$$Var[p,z]^{-1} = \begin{pmatrix} Var[p] & -\lambda_p \psi_I^{-1} \sigma_z^2 \\ -\lambda_p \psi_I^{-1} \sigma_z^2 & \sigma_z^2 \end{pmatrix}^{-1}$$
$$= \frac{1}{\lambda_p^2 Var[\hat{y}_I]\sigma_z^2} \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \\ \lambda_p \psi_I^{-1} \sigma_z^2 & Var[p] \end{pmatrix}$$

and covariance

$$Cov\left[y-p, \begin{pmatrix} p & z \end{pmatrix}\right] = \begin{pmatrix} Cov[y,p] - Var[p] & \lambda_p \psi_I^{-1} \sigma_z^2 \end{pmatrix}$$

Coefficient on price

$$\frac{(Cov[y,p] - Var[p])\sigma_z^2 + \lambda_p^2 \psi_I^{-2} \sigma_z^4}{\lambda_p^2 Var[\hat{y}_I]\sigma_z^2}(p - E[p]) = \frac{Cov[y,p] - \lambda_p^2 Var[\hat{y}_I] - \lambda_p^2 \psi_I^{-2}\sigma_z^2 + \lambda_p^2 \psi_I^{-2}\sigma_z^2}{\lambda_p^2 Var[\hat{y}_I]}(p - E[p])$$
$$= \left(\frac{\lambda_p Cov[y,\hat{y}_I]}{\lambda_p^2 Var[\hat{y}_I]} - 1\right)(p - E[p]) = \frac{Cov[y,\hat{y}_I] - \lambda_p Var[\hat{y}_I]}{\lambda_p Var[\hat{y}_I]}(p - E[p]).$$

and on stochastic supply/negative liquidity demand

$$\frac{\lambda_p \psi_I^{-1} \sigma_z^2}{\lambda_p^2 Var[\hat{y}_I]\sigma_z^2}\{Cov[y,p] - Var[p] + Var[p]\}\, z = -\frac{Cov[y,p]}{\lambda_p Var[\hat{y}_I]}\psi_I^{-1}(-z)$$

in baseline model $Cov[y,\hat{y}_I] = Var[\hat{y}_I]$, so $Cov[y,\hat{y}_I] - \lambda_p Var[\hat{y}_I] = (1-\lambda_p)Var[\hat{y}_I]$.

C.1.1. Noisy factor loadings

Including noisy factor loadings. Let $Var[p,z] = \boldsymbol{\Sigma}_{pz}$ and

$$\begin{aligned}
\boldsymbol{\Sigma}_{pz|\hat{\beta}} &= \boldsymbol{\Sigma}_{pz} - \begin{pmatrix} \lambda_p \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \\ \mathbf{0}^\top \end{pmatrix} \boldsymbol{D}_{\sigma_\beta}^{-2} \begin{pmatrix} \lambda_p \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q & \mathbf{0} \end{pmatrix} \\
&= \boldsymbol{\Sigma}_{pz} - \begin{pmatrix} \lambda_p \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{D}_{\sigma_\beta}^{-2} \\ \mathbf{0}^\top \end{pmatrix} \begin{pmatrix} \lambda_p \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q & \mathbf{0} \end{pmatrix} = \boldsymbol{\Sigma}_{pz} - \begin{pmatrix} \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q & 0 \\ 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} Var[p] - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q & -\lambda_p \psi_I^{-1} \sigma_z^2 \\ -\lambda_p \psi_I^{-1} \sigma_z^2 & \sigma_z^2 \end{pmatrix}
\end{aligned}$$

so

$$\begin{aligned}
|\boldsymbol{\Sigma}_{pz|\hat{\beta}}| &= (Var[p] - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q)\sigma_z^2 - \lambda_p^2 \psi_I^{-2} \sigma_z^4 \\
&= \lambda_p^2 Var[\hat{y}_I]\sigma_z^2 + \lambda_p^2 \psi_I^{-2} \sigma_z^4 - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q \sigma_z^2 - \lambda_p^2 \psi_I^{-2} \sigma_z^4 \\
&= \lambda_p^2 (Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q)\sigma_z^2
\end{aligned}$$

and

$$\boldsymbol{\Sigma}_{pz|\hat{\beta}}^{-1} = \frac{1}{\lambda_p^2 (Var[\hat{y}_I] - \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q)\sigma_z^2} \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \\ \lambda_p \psi_I^{-1} \sigma_z^2 & Var[p] - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q. \end{pmatrix}$$

Notice that

$$\begin{aligned}
-\boldsymbol{\Sigma}_{pz|\hat{\beta}}^{-1} \begin{pmatrix} \lambda_p \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \\ \mathbf{0}^\top \end{pmatrix} \boldsymbol{D}_{\sigma_\beta}^{-2} &= -|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \\ \lambda_p \psi_I^{-1} \sigma_z^2 & Var[p] - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q \end{pmatrix} \begin{pmatrix} \lambda_p \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \\ \mathbf{0}^\top \end{pmatrix} \boldsymbol{D}_{\sigma_\beta}^{-2} \\
&= -|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \begin{pmatrix} \sigma_z^2 \\ \lambda_p \psi_I^{-1} \sigma_z^2 \end{pmatrix} \lambda_p \boldsymbol{\mu}_q^\top,
\end{aligned}$$

and

$$\begin{aligned}
-\boldsymbol{D}_{\sigma_\beta}^{-2} \begin{pmatrix} \lambda_p \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q & \mathbf{0} \end{pmatrix} \boldsymbol{\Sigma}_{pz|\hat{\beta}}^{-1} &= -|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \boldsymbol{D}_{\sigma_\beta}^{-2} \begin{pmatrix} \lambda_p \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q & \mathbf{0} \end{pmatrix} \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \\ \lambda_p \psi_I^{-1} \sigma_z^2 & Var[p] - \lambda_p^2 \boldsymbol{\mu}_q^\top \boldsymbol{D}_{\sigma_\beta}^2 \boldsymbol{\mu}_q \end{pmatrix} \\
&= -|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1} \lambda_p \boldsymbol{\mu}_q \begin{pmatrix} \sigma_z^2 & \lambda_p \psi_I^{-1} \sigma_z^2 \end{pmatrix},
\end{aligned}$$

as well as

$$|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1}\begin{pmatrix}\lambda_p\boldsymbol{\mu}_q & \mathbf{0}\end{pmatrix}\begin{pmatrix}\sigma_z^2\\ \lambda_p\psi_I^{-1}\sigma_z^2\end{pmatrix}\lambda_p\boldsymbol{\mu}_q^\top = |\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1}\lambda_p^2\boldsymbol{\mu}_q\boldsymbol{\mu}_q^\top\sigma_z^2.$$

Moments: variance

$$\begin{aligned}
Var\left[p,z,\hat{\boldsymbol{\beta}}\right]^{-1} &= \begin{pmatrix}\boldsymbol{\Sigma}_{pz} & \begin{pmatrix}\lambda_p\boldsymbol{\mu}_q^\top\boldsymbol{D}_{\sigma_\beta}^2\\ \mathbf{0}^\top\end{pmatrix}\\ \begin{pmatrix}\lambda_p\boldsymbol{D}_{\sigma_\beta}^2\boldsymbol{\mu}_q & \mathbf{0}\end{pmatrix} & \boldsymbol{D}_{\sigma_\beta}^2\end{pmatrix}^{-1}\\
&= \begin{pmatrix}0\boldsymbol{I}_{2\times 2} & 0\boldsymbol{I}_{2\times n}\\ 0\boldsymbol{I}_{n\times 2} & \boldsymbol{D}_{\sigma_\beta}^{-2}\end{pmatrix} + |\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1}\begin{pmatrix}\boldsymbol{\Sigma}_{pz|\hat{\beta}}^{\text{adj}} & -\lambda_p\begin{pmatrix}\sigma_z^2\\ \lambda_p\psi_I^{-1}\sigma_z^2\end{pmatrix}\boldsymbol{\mu}_q^\top\\ -\lambda_p\boldsymbol{\mu}_q\begin{pmatrix}\sigma_z^2 & \lambda_p\psi_I^{-1}\sigma_z^2\end{pmatrix} & \lambda_p^2\boldsymbol{\mu}_q\boldsymbol{\mu}_q^\top\sigma_z^2\end{pmatrix}
\end{aligned}$$

covariance

$$\begin{aligned}
Cov\left[r,\begin{pmatrix}p & z & \hat{\boldsymbol{\beta}}\end{pmatrix}\right] &= \begin{pmatrix}Cov\left[r,\begin{pmatrix}p & z\end{pmatrix}\right] & -\lambda_p\boldsymbol{\mu}_q^\top\boldsymbol{D}_{\sigma_\beta}^2\end{pmatrix}\\
&= \begin{pmatrix}\lambda_p(Cov[y,\hat{y}_I]-\lambda_p Var[s_U]) & \lambda_p\psi_I^{-1}\sigma_z^2 & -\lambda_p\boldsymbol{\mu}_q^\top\boldsymbol{D}_{\sigma_\beta}^2\end{pmatrix} -
\end{aligned}$$

Coefficients

$$\begin{aligned}
&Cov\left[r,\begin{pmatrix}p & z & \hat{\boldsymbol{\beta}}\end{pmatrix}\right]Var\left[p,z,\hat{\boldsymbol{\beta}}\right]^{-1}\\
&\quad= \begin{pmatrix}\mathbf{0}^\top & -\boldsymbol{\mu}_q^\top\end{pmatrix}\\
&\qquad+|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1}\Bigg(Cov\left[r,\begin{pmatrix}p & z\end{pmatrix}\right]\boldsymbol{\Sigma}_{pz|\hat{\beta}}^{\text{adj}} + \lambda_p^2\boldsymbol{\mu}_q^\top\boldsymbol{D}_{\sigma_\beta}^2\boldsymbol{\mu}_q\begin{pmatrix}\sigma_z^2 & \lambda_p\psi_I^{-1}\sigma_z^2\end{pmatrix}\\
&\qquad\qquad - \lambda_p Cov\left[r,\begin{pmatrix}p & z\end{pmatrix}\right]\begin{pmatrix}\sigma_z^2\\ \lambda_p\psi_I^{-1}\sigma_z^2\end{pmatrix}\boldsymbol{\mu}_q^\top - \lambda_p^2\boldsymbol{\mu}_q^\top\boldsymbol{D}_{\sigma_\beta}^2\boldsymbol{\mu}_q\boldsymbol{\mu}_q^\top\sigma_z^2\Bigg)
\end{aligned}$$

where

$$
\begin{aligned}
&|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1}\left(Cov\left[r,\begin{pmatrix}p & z\end{pmatrix}\right]\boldsymbol{\Sigma}^{\mathrm{adj}}_{pz|\hat{\beta}}+\lambda_p^2\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q\begin{pmatrix}\sigma_z^2 & \lambda_p\psi_I^{-1}\sigma_z^2\end{pmatrix}\right)\begin{pmatrix}p-E[p]\\ z\end{pmatrix}\\
&=|\boldsymbol{\Sigma}_{pz|\hat{\beta}}|^{-1}\Bigg(\begin{pmatrix}\lambda_p(Cov[y,\hat{y}_I]-\lambda_p Var[s_U]) & \lambda_p\psi_I^{-1}\sigma_z^2\end{pmatrix}\begin{pmatrix}\sigma_z^2 & \lambda_p\psi_I^{-1}\sigma_z^2\\ \lambda_p\psi_I^{-1}\sigma_z^2 & Var[p]-\lambda_p^2\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q\end{pmatrix}\\
&+\lambda_p^2\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q\begin{pmatrix}\sigma_z^2 & \lambda_p\psi_I^{-1}\sigma_z^2\end{pmatrix}\Bigg)\begin{pmatrix}p-E[p]\\ z\end{pmatrix}\\
&=\frac{Cov[y,\hat{y}_I]-\lambda_p(Var[\hat{y}_I]+\psi_I^{-1}\sigma_z^2)+\lambda_p\psi_I^{-1}\sigma_z^2+\lambda_p\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q}{\lambda_p^2(Var[\hat{y}_I]-\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q)\sigma_z^2}\lambda_p\sigma_z^2(p-E[p])\\
&\quad+\lambda_p\psi_I^{-1}\sigma_z^2\frac{\lambda_p Cov[y,\hat{y}_I]-Var[p]+Var[p]-\lambda_p^2\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q+\lambda_p^2\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q}{\lambda_p^2(Var[\hat{y}_I]-\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q)\sigma_z^2}z\\
&=\frac{Cov[y,\hat{y}_I]-\lambda_p(Var[\hat{y}_I]-\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q)}{\lambda_p(Var[\hat{y}_I]-\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q)}(p-E[p])-\frac{Cov[y,\hat{y}_I]}{\lambda_p(Var[\hat{y}_I]-\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q)}(-\psi_I^{-1}z)
\end{aligned}
$$

and

$$
\begin{aligned}
&-\left\{1+\sigma_z^2\frac{\lambda_p Cov[y,\hat{y}_I]-\lambda_p^2 Var[s_U]+\lambda_p^2\psi_I^{-2}\sigma_z^2+\lambda_p^2\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q}{\lambda_p^2(Var[\hat{y}_I]-\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q)\sigma_z^2}\right\}\boldsymbol{\mu}_q^\top(\hat{\boldsymbol{\beta}}-\boldsymbol{\mu}_\beta)\\
&\quad=-\frac{Cov[y,\hat{y}_I]}{\lambda_p(Var[\hat{y}_I]-\boldsymbol{\mu}_q^\top\boldsymbol{D}^2_{\sigma_\beta}\boldsymbol{\mu}_q)}\boldsymbol{\mu}_q^\top(\hat{\boldsymbol{\beta}}-\boldsymbol{\mu}_\beta).
\end{aligned}
$$

C.2. Trading volume

Realized trading volume v is given by

$$
\begin{aligned}
v&=\frac{1}{2}\{|\delta_I|+|\delta_U|+|z|\}\\
&=\frac{1}{2}\Big\{\psi_I\left|(1-\lambda_p)(\hat{y}_I-E[\hat{y}_I])-\lambda_p\psi_I^{-1}z\right|+\psi_U(\lambda_p-\lambda_U)\left|(E[\hat{y}_I]-\hat{y}_I-\psi_I^{-1}z)\right|+|z|\Big\}\\
&=\frac{1}{2}\Big\{\psi_I\left|(1-\lambda_p)(\hat{y}_I-E[\hat{y}_I])-\lambda_p\psi_I^{-1}z\right|+\psi_I(1-\lambda_p)\left|-(\hat{y}_I-E[\hat{y}_I])-\psi_I^{-1}z\right|+|z|\Big\}\\
&=\frac{\psi_I}{2}\Big\{\left|(1-\lambda_p)(\hat{y}_I-E[\hat{y}_I])-\lambda_p\psi_I^{-1}z\right|+(1-\lambda_p)\left|(\hat{y}_I-E[\hat{y}_I])+\psi_I^{-1}z\right|+\psi_I^{-1}|z|\Big\},
\end{aligned}
$$

which in the third line uses the equality

$$
\psi_U(\lambda_p-\lambda_U)=\psi_U\frac{\psi_I+\lambda_U\psi_U-\lambda_U(\psi_I+\psi_U)}{\psi_I+\psi_U}=\psi_I\frac{(1-\lambda_U)\psi_U}{\psi_I+\psi_U}=\psi_I(1-\lambda_p).
$$

C.3. Expected profit and ex-ante expected utility

C.3.1. Expected profit

Using that

$$
\begin{aligned}
&\hat{y}_I - p = (1-\lambda_p)\hat{y}_I - (1-\lambda_p)\boldsymbol{\mu}_\beta^\top \boldsymbol{\mu}_q - \lambda_p \psi_I^{-1} z = (1-\lambda_p)(\hat{y}_I - E[\hat{y}_I]) - \lambda_p \psi_I^{-1} z \\
&\implies E[\hat{y}_I - p] = 0 \\
&\implies E[(\hat{y}_I - p)^2] = Var[\hat{y}_I - p] = (1-\lambda_p)^2 Var[\hat{y}_I] + \lambda_p^2 \psi_I^{-2}\sigma_z^2
\end{aligned}
$$

key expectations are

$$
\begin{aligned}
E[(\hat{y}_I - p)y] &= (1-\lambda_p)Cov[y, \hat{y}_I] = (1-\lambda_p)\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\mu}_\beta \\
E[(\hat{y}_I - p)p] &= Cov[\hat{y}_I, p] - Var[p] = \lambda_p Var[\hat{y}_I] - \lambda_p^2 \left\{ Var[\hat{y}_I] + \psi_I^{-2}\sigma_z^2 \right\} \\
&= \lambda_p(1-\lambda_p)Var[\hat{y}_I] - \lambda_p^2 \psi_I^{-2}\sigma_z^2
\end{aligned}
$$

and informed profit under the objective measure is

$$
\begin{aligned}
E[\pi_I] &= \psi_I E\left[(\hat{y}_I - p)(y-p)\right] = \psi_I \left\{ E[(\hat{y}_I - p)y] - E[(\hat{y}_I - p)p] \right\} \\
&= \psi_I \left\{ (1-\lambda_p)\left(Cov[y, \hat{y}_I] - \lambda_p Var[\hat{y}_I] \right) + \lambda_p^2 \psi_I^{-2}\sigma_z^2 \right\}.
\end{aligned}
$$

Under the contemporary measure $E[\boldsymbol{\beta}|\boldsymbol{c}^*] = \boldsymbol{\mu}_\beta$, covariance is the quadratic form $Cov[y, \hat{y}_I|\boldsymbol{c}^*] = \boldsymbol{\mu}_\beta^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\mu}_\beta$, profits are

$$
E[\pi_I|\boldsymbol{c}^*] = \psi_I \left\{ (1-\lambda_p)\left(Cov[y, \hat{y}_I|\boldsymbol{c}^*] - \lambda_p Var[\hat{y}_I] \right) + \lambda_p^2 \psi_I^{-2}\sigma_z^2 \right\},
$$

and out of sample surprise is

$$
E[\pi_I] - E[\pi_I|\boldsymbol{c}^*] = \psi_I(1-\lambda_p)(Cov[y, \hat{y}_I] - Cov[y, \hat{y}_I|\boldsymbol{c}^*]).
$$

Similar analysis for the uninformed profit yields

$$
\begin{aligned}
\hat{y}_U - p = (\lambda_p - \lambda_U)(E[\hat{y}_I] - s_U) &\implies E[\hat{y}_U - p] = 0 \\
&\implies E[(\hat{y}_U - p)^2] = (\lambda_p - \lambda_U)^2 Var[s_U]
\end{aligned}
$$

so

$$
E[(\hat{y}_U - p)y] = -(\lambda_p - \lambda_U)Cov[y, \hat{y}_I], \quad E[(\hat{y}_U - p)p] = -(\lambda_p - \lambda_U)\lambda_p Var[s_U]
$$

and expected profit under the objective measure is

$$\begin{aligned}E[\pi_U] &= \psi_U E\left[(\hat{y}_U - p)(y-p)\right] = \psi_U \left\{E[(\hat{y}_U - p)y] - E[(\hat{y}_U - p)p]\right\} \\ &= \psi_U \left\{(\lambda_p - \lambda_U)\left(\lambda_p Var[\hat{y}_I] - Cov[y, \hat{y}_I]\right) + (\lambda_p - \lambda_U)\lambda_p \psi_I^{-2}\sigma_z^2\right\}.\end{aligned}$$

Notice that

$$\psi_U(\lambda_p - \lambda_U) = \psi_U \frac{\psi_I + \lambda_U \psi_U - \lambda_U(\psi_I + \psi_U)}{\psi_I + \psi_U} = \psi_I \frac{(1-\lambda_U)\psi_U}{\psi_I + \psi_U} = \psi_I(1-\lambda_p)$$

so out of sample surprise is

$$E[\pi_U] - E[\pi_U|\boldsymbol{c}^*] = \psi_I(1-\lambda_p)(Cov[y, \hat{y}_I|\boldsymbol{c}^*] - Cov[y, \hat{y}_I]).$$

C.3.2. Difference in profit under contemporaneous measure

Expected difference in profit under the contemporaneous measure is

$$\begin{aligned}E[\pi_I|\boldsymbol{c}^*] - E[\pi_U|\boldsymbol{c}^*] &= \psi_I \left\{(1-\lambda_p)\left(Cov[y, \hat{y}_I|\boldsymbol{c}^*] - \lambda_p Var[\hat{y}_I]\right) + \lambda_p^2 \psi_I^{-2}\sigma_z^2\right\} \\ &\quad - \psi_I(1-\lambda_p)\left\{\left(\lambda_p Var[\hat{y}_I] - Cov[y, \hat{y}_I|\boldsymbol{c}^*]\right) + \lambda_p \psi_I^{-2}\sigma_z^2\right\} \\ &= \psi_I \left\{2(1-\lambda_p)\left(Cov[y, \hat{y}_I|\boldsymbol{c}^*] - \lambda_p Var[\hat{y}_I]\right) + \left[2\lambda_p - 1\right]\lambda_p \psi_I^{-2}\sigma_z^2\right\}\end{aligned}$$

A necessary condition for the differential to be positive is that the last term in the curly bracket, $[2\lambda_p - 1]\,\lambda_p \psi_I^{-2}\sigma_z^2$, which require that informed investors trade more aggresively $\psi_I > \psi_U$, since

$$2\lambda_p > 1 \iff 2\psi_I + 2\lambda_U \psi_U > \psi_I + \psi_U \iff \psi_I + (2\lambda_U - 1)\psi_U > 0 \impliedby \psi_I > \psi_U.$$

For symmetric uncertainty aversion, this simplifies to the informed making better prediction (see Proposition 3).

C.3.3. Ex-post performance

Ex-post performance surprises are symmetric and only exist with non-zero bias

$$\begin{aligned}&E[\pi_I] - E[\pi_I|\boldsymbol{c}^*] = \psi_I(1-\lambda_p)\left\{Cov[y, \hat{y}] - Cov[y, \hat{y}|\boldsymbol{c}^*]\right\} = \psi_I(1-\lambda_p)\left\{\boldsymbol{\varepsilon}_\beta \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta - \boldsymbol{\beta}\boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta\right\}, \\ &\text{and} \quad E[\pi_U] - E[\pi_U|\boldsymbol{c}^*] = -\psi_I(1-\lambda_p)\left\{\boldsymbol{\varepsilon}_\beta \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta - \boldsymbol{\beta}\boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta\right\}.\end{aligned}$$

By algebraic manipulation

$$Cov[y,\hat{y}] - Cov[y,\hat{y}|\boldsymbol{c}^*] = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\beta^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\mu}_\beta = \boldsymbol{\varepsilon}_\beta \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta - \boldsymbol{\beta} \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta$$

and

$$\psi_U(\lambda_p - \lambda_U) = \psi_U \frac{\psi_I + \lambda_U \psi_U - \lambda_U(\psi_I + \psi_U)}{\psi_I + \psi_U} = \psi_I \frac{(1-\lambda_U)\psi_U}{\psi_I + \psi_U} = \psi_I(1-\lambda_p).$$

Of the two components of the cost of complexity, ex-post performance surprises are entirely driven by the bias, and while the sign of the first term in the curly bracket could be both negative or positive, the quadratic form is always positive due to the positive definiteness of $\boldsymbol{\Sigma}_\zeta$, suggests that the term might be positive more often than not.

Ex-post performance surprises are a transfer between investors, and due to their symmetry it leaves certain results from the baseline model unaltered regardless of its sign. Ex-post performance surprises are a transfer between investors and nets out in total

$$E[\pi_I] - E[\pi_I|\boldsymbol{c}^*] - (E[\pi_U] - E[\pi_U|\boldsymbol{c}^*]) = 0.$$

It follows that for matters concerning the total profits of investors the distinction between objective measure and contemporary measure is irrelevant. The corresponding result in the baseline model arises trivially because $Var[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]] = Cov[y, E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]]$ so there are no ex-post surprises. Due to the common component in expected profit, total profit of investors simplifies to

$$E[\pi_I] + E[\pi_U] = E[\pi_I|\boldsymbol{c}^*] + E[\pi_U|\boldsymbol{c}^*] = \psi_I^{-1}\sigma_z^2 = \alpha_I E[(y-\hat{y}_I)^2]\sigma_z^2.$$

This result mirrors a result in the baseline model where it holds with the modification that the mean squared error collapses to the conditional variance under the true model. Introducing noisy estimation does not alter the intuition of the baseline model that total profits are increasing in the quality of predictions made by informed investors.

Appendix D. Value of data

D.1. Rational function formulation of cost of complexity

Let the adjacency matrix of any matrix $\boldsymbol{A}$ be indicated by superscript $\boldsymbol{A}^{\text{adj}}$ and the determinant by $|\boldsymbol{A}|$, and define

$$\boldsymbol{W} := k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta} = k_c^{-2}\boldsymbol{\Omega}_{\zeta 0} + \boldsymbol{D}_{\Omega_\zeta 0} + k_S(k_c^{-2}\boldsymbol{S} + \boldsymbol{D}_S) := \boldsymbol{W}_0 + k_S\boldsymbol{W}_S$$

such that

$$\begin{aligned}
\mathbf{1}^\top\boldsymbol{X}^{-1}\mathbf{1} &= \mathbf{1}^\top\left\{k_c^2\boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1}\right\}^{-1}\mathbf{1} = \mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta}\mathbf{1} - \mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta}\left\{k_c^{-2}\boldsymbol{\Omega}_\zeta + \boldsymbol{D}_{\Omega_\zeta}\right\}^{-1}\boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \\
&= \mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta}\mathbf{1} - \mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{W}^{-1}\boldsymbol{D}_{\Omega_\zeta} = \boldsymbol{D}_{\Omega_\zeta}\mathbf{1} - |\boldsymbol{W}|^{-1}\mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta}\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \\
&= \mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta}\mathbf{1} \\
&\quad - |\boldsymbol{W}|^{-1}\left\{\mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta 0}\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_{\Omega_\zeta 0}\mathbf{1} + k_S^2\mathbf{1}^\top\boldsymbol{D}_S\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_S\mathbf{1} + 2k_S\mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta 0}\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_S\mathbf{1}\right\}.
\end{aligned}$$

For a sum of matrices where one is scaled by a scalar k such as $\boldsymbol{W}$, it can be shown, see Appendix D.2, that the determinant is a polynomial in k of degree equal to the number of rows (or columns) of the matrix, and that the sum over its adjugate matrix is a polynomial in k of one degree less. Scaling the entries of the adjugate matrix before taking the sum, as is done in the terms of the curly bracket above, does not change the degree of the resulting polynomial. However, the multiplication of the second term by the square of k_S, which is the variable of the polynomial for $\boldsymbol{W}$, yields a polynomial of a degree one *higher* than the number of rows, which for $\boldsymbol{W}$ is equal to the number of factor n_q. The curly bracket divided by the determinant is therefore a rational function (ratio of polynomials) of k_S. By extension, the cost of complexity is a rational function of k_S

$$\begin{aligned}
\chi(k_S) = k_{\sigma 0}^2|\boldsymbol{W}|^{-1}\Big\{&|\boldsymbol{W}|\mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta 0}\mathbf{1} + k_S|\boldsymbol{W}|\mathbf{1}^\top\boldsymbol{D}_S\mathbf{1} - \mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta 0}\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_{\Omega_\zeta 0}\mathbf{1} \\
&- k_S^2\mathbf{1}^\top\boldsymbol{D}_S\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_S\mathbf{1} - 2k_S\mathbf{1}^\top\boldsymbol{D}_{\Omega_\zeta 0}\boldsymbol{W}^{\text{adj}}\boldsymbol{D}_S\mathbf{1}\Big\} = k_{\sigma 0}^2\frac{\sum_{\ell=0}^{n_q+1}a_\ell k_S^\ell}{\sum_{\ell=0}^{n_q}b_\ell k_S^\ell}.
\end{aligned}$$

The degree of the polynomial in the numerator is exactly one higher than the degree of the polynomial in the denominator and the rational function therefore has a an oblique asymptote, which is linear in k_S. The slope of the asymptote is the coefficient of the highest power of the polynomial in the numerator divided by the coefficient of the highest power in the denominator a_{n_q+1}/b_{n_q}, or the limit of the derivative with respect to k_S, i.e. $\lim_{k_S\to\infty}\partial\chi(k_S)/\partial k_S$.

A special case is that of a diagonal $\boldsymbol{S}$ matrix, in which case the derivative simplifies to

$$\lim_{k_S \to \infty} \frac{\partial \chi(k_S)}{\partial k_S} = \frac{k_{\sigma 0}^2}{1 + k_c^2} \mathbf{1}^\top \boldsymbol{D}_S \mathbf{1},$$

and in the even more special case where all true factor loadings are the same $\boldsymbol{\beta} = \bar{\beta}\mathbf{1}$, whether more data (stronger signal) is asymptotically valuable or value destroying is entirely determined by the base-parameters $k_{\sigma 0}$, k_c and $\bar{\beta}$, specifically

$$\lim_{k_S \to \infty} \frac{\partial \chi(k_S)}{\partial k_S} + \frac{\partial Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I]}{\partial k_S} = \mathbf{1}^\top \boldsymbol{D}_S \mathbf{1} \left\{ \frac{k_{\sigma 0}^2}{1 + k_c^2} - \bar{\beta}^2 \right\}.$$

D.2. Matrix inverses as rational functions

If a square matrix $\boldsymbol{L} = \boldsymbol{A} + k\boldsymbol{B}$ is of dimensions two by two matrix, its determinant is a polynomial in k of the form

$$|\boldsymbol{L}_{2\times 2}| = |\boldsymbol{A}| + k^2|\boldsymbol{B}| + k(a_{11}b_{22} + a_{22}b_{11} - a_{12}b_{21} - a_{21}b_{12}).$$

Let $\boldsymbol{M}_{ij}$ be the $(n-1) \times (n-1)$ sub-matrix of the square $n \times n$ matrix $\boldsymbol{L}$ that deletes row i and column j, i.e. its determinant $|\boldsymbol{M}_{ij}|$ is the ij-minor of $\boldsymbol{L}$. By a Laplace expansion over row i, the determinant of $\boldsymbol{L}$ is given by $|\boldsymbol{L}| = \sum_j (-1)^{i+j} l_{ij} |\boldsymbol{M}_{ij}|$, where the entries of $\boldsymbol{L}$ are of the form $l_{ij} = a_{ij} + kb_{ij}$. Therefore, if $|\boldsymbol{M}_{ij}|$ is a polynomial in k, the product $l_{ij}|\boldsymbol{M}_{ij}|$ is a polynomial in k of one degree higher. Since $|\boldsymbol{L}_{2\times 2}|$ is a polynomial of second degree, the determinant of $\boldsymbol{L}_{n\times n}$ is a polynomial of n'th degree. Meanwhile, the co-factor matrix of $\boldsymbol{L}$ denoted $\boldsymbol{C}$ has entries $c_{ij} = (-1)^{i+j} l_{ij} |\boldsymbol{M}_{ij}|$ which are polynomials in k of degree $n-1$. The sum over the adjugate matrix $\mathbf{1}^\top \boldsymbol{L}^{\mathrm{adj}} \mathbf{1} = \mathbf{1}^\top \boldsymbol{C}^\top \mathbf{1}$ is, therefore, a polynomial of degree $n-1$.