Equilibrium Data Mining and Data Abundance*

Jérôme Dugast† Thierry Foucault‡

First Version: November 2019.
This version: April 8, 2020

WORK IN PROGRESS-COMMENTS WELCOME.

Abstract

We analyze how information processing power and data abundance affect speculators' search for predictors. Speculators optimally search for a predictor whose signal-to-noise ratio exceeds an endogenous threshold. Greater computing power raises this threshold, and therefore price informativeness, because it reduces the cost of search. In contrast, data abundance can lower this threshold because (i) it intensifies competition among speculators, which reduces the benefit of finding a good predictor and (ii) it increases the total expected cost of finding a predictor. In the former (latter) case, price informativeness increases (decreases) with data abundance. We present additional testable implications of these effects.

Keywords: Alternative Data, Data Abundance, Data Mining, Price Informativeness, Search for Information.

*We are grateful to Bruno Biais, Jean-Edouard Colliard, Johan Hombert, Denis Gromb, Daniel Schmidt and seminar participants at HEC Paris for very useful comments. Future versions will be available at: https://sites.google.com/view/jeromedugast/home or https://thierryfoucault.com/.

†Université Paris-Dauphine, PSL. Tel: (+33) 01 44 05 40 41 ; E-mail: jerome.dugast@dauphine.psl.eu

‡HEC, Paris and CEPR. Tel: (33) 1 39 67 95 69; E-mail: foucault@hec.fr

1. Introduction

Asset managers devote considerable effort to find new investment signals (predictors). To this end, they have made massive investments to leverage progress in information technologies, namely the considerable growth in available data (data abundance) and the steep decline in the cost of data processing, due to progress in computing power (see Nordhaus (2015)). For instance, asset managers increasingly buy so called "alternative data sets", including credit/debit card data, app usage data, satellite images, social media, web traffic data, etc. and use computer-based methods to extract predictors of asset payoffs from these data.[1]

Data abundance and improvements in computing power are related but distinct phenomena. For instance, satellite images or web traffic data can provide more accurate predictors of future firms' earnings.[2] However, they do not per se reduce the cost of processing data for obtaining these predictors. Thus, understanding the effects of data abundance require models of information acquisition in which the effects of an increase in the amount of available data can be analyzed holding the cost of data processing constant (and vice versa). In this paper, we propose a model of this type and we show that the effects of data abundance and progress in computing power on equilibrium outcomes in financial markets (e.g., the dispersion of speculators' trading profits and price informativeness) are distinct.

Our model features a continuum of risk averse speculators. In the first stage (the "exploration stage"), each speculator optimally scours available data to find a predictor of the payoff of a risky asset. In the second stage (the "trading stage"), each speculator observes the realization of her predictor and optimally chooses her trading strategy. We formalize the trading stage as a standard rational expectations model (similar to Vives (1995)). The novelty of our model (and its implications) stems from the exploration stage. Here, instead of following the standard approach (e.g., Grossman and Stiglitz (1980) or Verrecchia (1982)), whereby speculators obtain a predictor of a given precision in exchange of a payment, we explicitly model the search for a predictor as a sequential process and we analyze how the optimal search strategy depends on (i) the cost of exploration and (ii) the amount of data available for exploration.

[1] Marenzi (2017) estimates that asset managers have spent more than 4 bio in alternative data in 2017 (see also "Asset managers double spending in new data in hunt for edge", Financial Times, May 9, 2018.. Abis (2018) finds that quantitative funds (using computer-driven models to analyze large datasets) have quadrupled in size from 1999 to 2015 and that their growth has been more than twofold that of discretionary funds. Moreover, Grennan and Michaely (2019) find that 57% of the FinTechs in their sample specialize in producing investment signals using artificial intelligence.

[2] See Katona et al. (2019) and Zhu (2019) for evidence that aggregated signals from alternative data (e.g., consumer browsing-data or satellite images) have predictive power for firms' future earnings.

We model the search for predictors as follows. Predictors differ in their signal-to-noise ratios ("quality"), whose distribution is exogenous. The amount of available data determines the quality of the most informative predictor (the data frontier), denoted τ^{max}. The least informative predictor is uninformative. Given this distribution, each speculator simultaneously and independently explores ("mines") the data. Each new exploration costs c and returns a predictor whose quality is drawn from the distribution of predictors' quality. After obtaining a predictor, a speculator can decide either to explore the data further, to possibly obtain an even better predictor, or to trade on her latest predictor.[3]

As a motivation for this modeling approach, consider asset managers using accounting variables to forecast future stock earnings. There are many ways to combine these variables to obtain predictors. For instance, using 240 accounting variables from firms' financial statements, Yan and Zheng (2017) build more than 18,000 trading signals and find that many of these signals can be used to build trading strategies with significant abnormal returns (even after accounting for the risk of data snooping). The data mining cost, c, can be viewed as the labor cost of considering a particular combination (a predictor), designing a trading strategy based on this predictor, backtesting it, and thinking about possible economic stories for why the strategy works. After obtaining a predictor, each manager can decide to start trading on it or to keep searching for another, more accurate, predictor.

New datasets enable speculators to use new variables to forecast asset payoffs and should therefore push back the data frontier, τ^{max}.[4] Thus, in the baseline version of our model, we study the effect of data abundance by analyzing how an increase in τ^{max} affects equilibrium outcomes. As for greater computing power, it reduces the cost of exploring a new dataset.[5] Thus, we study the effect of greater computing power by considering the effect of a decrease in c on equilibrium outcomes.

In equilibrium, each speculator's optimal search strategy follows a stopping rule: She stops

[3] In our model, after finding a predictor, speculators perfectly observe its quality. Thus, we focus on how speculators optimally mine the data, not on the risk of false discoveries associated with data mining, as is often emphasized by financial economists (see Harvey (2017)).

[4] For instance, Katona et al. (2019) find that combining satellite images of parking lots of U.S. retailers from two distinct data providers improve the accuracy of the forecasts of these firms' quartely earnings based on these images.

[5] For instance, an increase in computing power reduces the time costs of finding predictors. Brogaard and Zareei (2019) use a genetic algorithm approach to select technical trading rules. They note that "*the average time needed to find the optimum trading rules for a diversified portfolio of ten NYSE/AMEX volatility assets for the 40 year sample using a computer with an Intel® Core(TM) CPU i7-2600 and 16 GM RAM is 459.29 days (11,022.97 hours)." For one year it takes approximately 11.48 days.*" They conclude that their analysis would not be possible without the considerable increase in computing power in the last 20 years.

searching for a predictor after finding one whose quality (signal-to-noise ratio) exceeds an endogenous threshold, denoted τ^* (we refer to such a predictor as being "satisficing"). This threshold is such that the speculator's expected utility of trading on a predictor of quality τ^* is just equal to her expected utility of searching for another predictor. The latter reflects the prospect of obtaining a larger expected trading profit by finding a predictor of higher quality deflated by the *total* expected cost of search to find such a predictor the per-exploration cost, c times the expected number of explorations required to find a predictor with a quality higher than τ^*). All speculators use the same stopping rule because they are ex-ante identical (same preferences, search cost etc.). However, as explorations' outcomes are random, speculators find and trade on predictors of different quality. Thus, in equilibrium, (i) only predictors of sufficiently high quality are used for trading and (ii) speculators endogenously exploit predictors of different quality. Specifically, the quality of predictors used in equilibrium ranges from τ^* (the least informative predictor used in equilibrium) to τ^{max} (most informative).

Greater computing power induces speculators to adopt a more stringent stopping rule in equilibrium, i.e., a decrease in c raises τ^*. Indeed, a decrease in the per-exploration cost, c, directly reduces the total expected cost of launching a new exploration after finding a predictor. Hence, it raises the value of searching for another predictor after finding one and therefore it induces speculators to be more demanding for the quality, τ^*, of the least informative predictor used in equilibrium. One indirect effect of this behavior is that, on average, speculators trade more aggressively on their signal (we call this the "competition effect") because they face less uncertainty on the asset payoff since their predictors are better on average. As a result, price informativeness increases. This indirect effect of a decrease in the per exploration cost dampens its direct positive effect on the value of searching for a better predictor after finding one. However, it is never strong enough to fully offset the decrease in the total expected search cost due to a decrease in c.

The effect of pushing back the data frontier (an increase in τ^{max}) on the optimal stopping rule is more subtle because it directly affects the value of searching for another predictor after finding one in two opposite directions. One the one hand, it raises this value for two reasons. First, holding the stopping rule constant, it enlarges the range of satisficing predictors, which raises the probability that each exploration is successful. This effect reduces the total expected cost of search. Second, holding price informativeness constant, it increases the expected utility of trading on a satisficing predictor due to the prospect of finding even more informative predic-

tors (the "hidden gold nugget effect"). However, an increase in the quality of the best predictor has also a direct positive effect on price informativeness because it raises the average quality of predictors and therefore the average aggressiveness with which speculators exploit their signals (intuitively, those with the most informative signals exploit them very aggressively). This competition effect reduces the value of searching for predictors. We show that it always dominates the other effects when τ^{max} is high enough. When this happens, a push back of the data frontier leads speculators to follow a less stringent stopping rule in their search for predictors. Thus, the model implies an inverse U-shape relationship between the quality of the least informative predictor used in equilibrium (τ^*) and data abundance.

In the baseline model, data abundance increases the likelihood of finding a satisficing predictor because it pushes back the data frontier while leaving unchanged the odds of finding no predictor. In reality, it can also lead to a "needle in the haystack" problem by making it more difficult to identify truly informative datasets. To account for this possibility, we extend our baseline model by assuming that each exploration returns an informative predictor with probability $\alpha < 1$ (in the baseline model, $\alpha = 1$) and analyze the "needle in the haystack" problem by considering the effects of a drop in α. Such a drop reduces the chance of finding a satisficing predictor in a given exploration and therefore it raises the total expected cost of search for speculators. For this reason, it leads speculators to be *less* demanding for the quality of the least informative predictor, τ^*, for the same reasons as an *increase* in the per exploration cost does.

In sum, the model highlights two channels through which data abundance can reduce the quality of the least informative predictor used in equilibrium: (i) It reduces the trading value of predictors by intensifying competition among speculators (the competition effect) and (ii) it increases the total expected cost of search, even though it does not change per exploration cost ("needle in the haystack effect").

The model suggests several ways to test the differing implications of the various facets of progress in information technologies (a reduction in c, an increase in τ^{max}, and a decrease in α). The most direct way maybe consists in estimating the quality of predictors used by speculators (e.g., fund managers). This could be done by running a time-series regression of a speculator's position in one asset on the return of this asset (e.g., using quarterly data). The coefficient of this regression is a measure of the speculator's stock picking ability and it is a proxy for the quality of her predictor. It could therefore be used to rank speculators according to their stock

picking ability. Our model predicts that improvements in computing power (a decrease in c) should increase the stock picking ability of speculators in the lowest rank (a proxy for τ^*) while data abundance (an increase in τ^{max}, or a decrease in α) can have the opposite effects.

Our second set of predictions is about asset price informativeness. Our model predicts that greater computing power improves price informativeness because it leads speculators to be more demanding for the quality of their predictors.[6] In contrast, the effect of data abundance on asset price informativeness is more complex. On the one hand, a push back the data frontier (as in the baseline version of the model) always improves price informativeness. Indeed, in this case, the average quality of predictors increases because, in equilibrium, the potential drop in the quality of the least informative predictor is always more than offset by the increase in the quality of the most informative predictor. In contrast, as the needle in the haystack problem becomes more acute (a drop in α), price informativeness drops because speculators adopt a less demanding stopping rule for their predictors and therefore the average quality of predictors drops.

These results suggest that the effects of progress in information technologies on asset price informativeness are likely to be ambiguous as this progress triggered both a growth in computing power and available data. Consistent with this implication, empirical studies find that the evolution of stock price informativeness over the long run has been ambiguous. For instance Bai et al. (2016) find that the price stocks in the SP500 has become more informative since 1960 while Farboodi et al. (2019) find the opposite patterns for all stocks, except for large growth stocks. Interestingly, Zhu (2019) finds that the introduction of satellite images and consumer transactions has a positive effect on stock price informativeness. This finding is consistent with our model as, arguably, access to satellite images and consumer transactions data should allow investors to build even more accurate predictors of future earnings (an increase in τ^{max}).

Our third set of predictions regards effects of computing power and data abundance on trading profits (excess returns) and holdings for speculators (e.g., mutual funds). The model predicts an inverse U-shape relationship between speculators' average trading profits and computing power. Indeed, greater computing power raises the average quality of the predictors used in equilibrium and therefore price informativeness. The first effect raises speculators' expected trading profit while the second reduces it. The former dominates if and only if c is large enough.

[6] In line with this prediction, Gao and Huang (2019) find that the introduction of the EDGAR system in the U.S. (which allows investors to have internet access to electronic filings by firms) had a positive effects on measures of price efficiency. One possible reason, as argued by Gao and Huang (2019), is that the EDGAR system reduced the cost of accessing data (a component of exploration cost) for investors.

An improvement in the data frontier has the same effect for the same reasons. In contrast, an intensification of the needle in the haystack problem operates in the opposite direction (there is U-shape relationship between α and speculators' expected profit) because its effects are similar to that of an increase in c. Moreover, an increase in computing power always reduces the dispersion of trading profits among speculators while data abundance can increase it. Finally, we show that greater computing power or an improvement in the data frontier reduce the pairwise correlation in speculators' trades (or the average pairwise correlation across all speculators) while a drop in the proportion of informative datasets (α) has the opposite effect.

Our paper contributes to the large literature on informed trading with endogenous information acquisition (e.g., Grossman and Stiglitz (1980), Verrecchia (1982); see Veldkamp (2011) for a survey). This literature often takes a reduced-form approach to model the cost of acquiring a signal of given precision. For instance, Verrecchia (1982) (and several authors) assumes that this cost is a convex function of the precision of the signal. The learning technology in our model is very different. Indeed, speculators do not directly choose the precision of their signal. Rather they partially control the distribution from which this precision is drawn through the choice of their stopping rule and effectively pay a larger total expected cost to obtain a higher precision on average (by choosing a more stringent stopping rule). Moreover, this specification is not assumed but micro-founded by an optimal search model.[7] As explained previously, this approach gives us a way to analyze separately the effects of greater computing power (a decrease in the cost of processing data) and data abundance.

Our paper is also related to the recent literature analyzing the economic effects of progress in information technologies (see, Veldkamp and Chung (2020) for a review) and more specifically theoretical papers analyzing the effects of these technologies for the production of financial information (e.g., Dugast and Foucault (2018), Farboodi and Veldkamp (2019), or Shyang Huang and Yang (2020)). These papers usually analyze this progress as a decrease in the cost of processing information. In contrast, our model focuses on another dimension of this progress, namely the abundance of new data. We show that the effects of these two dimensions on the

[7]Han and Sangiorgi (2018) considers a model in which agents can choose the number of signals about a variable drawn, with replacement, from an urn. Each draw is costly in their model. They show that this way of modeling information acquisition provides a micro-foundation for set-ups in which (i) agents receive a signal with an additive common and idiosyncratic noise, (ii) the precision of each agent signal is increasing and concave in the amount invested but (iii) the precision of the idiosyncratic component of each signal is increasing and convex signal in this amount. Their model of information acquisition does not allow for sequential search (the decision to draw another signal cannot be contingent on the outcome of past draws) and they do not relate agents' expected payoff to their trading decision in the market for an asset.

incentive to produce financial information are distinct and we derive several implications that should allow empiricists to test whether this distinction is important empirically.

2. Model

We consider the market for a risky asset populated with a unit mass continuum of risk averse (CARA) speculators, a risk neutral and competitive market maker, and noise traders. The payoff of the asset, ω, is realized in period 2 and is normally distributed with mean zero and variance σ^2. Speculators search for predictors of the asset payoff in period 0 (the "exploration stage"). Then, in period 1 (the "trading stage"), they observe the realization of these predictors and can trade on them in the market for the risky asset. We now describe these two stages in details.

The exploration stage. In period 0, each speculator i searches for a *predictor* of the asset payoff, ω. There is a continuum of potential predictors. A predictor s_θ is characterized by its type θ and is such that:

$$s_\theta = \cos(\theta)\omega + \sin(\theta)\varepsilon_\theta, \tag{1}$$

where $\theta \in [0, \pi/2]$ and ε_θs are normally and independently distributed with mean zero and variance σ^2. Moreover, ε_θ is independent from ω. Let $\tau_\theta \equiv (\frac{\cos^2(\theta)}{\sin^2(\theta)}) = \cot^2(\theta)$ denote the signal-to-noise ratio for a predictor with type θ. We refer to this ratio as the "quality" of a predictor.[8] The quality of a predictor is inversely related to its type, θ and unrelated to the risk of the asset, σ^2, because $\text{Var}(\varepsilon_\theta) = \text{Var}(\omega) = \sigma^2$. Without this assumption, an increase in the risk of the asset would artificially increase the quality of all predictors.

We assume that predictors's types, θ, are distributed according to the density $\phi(.)$ on $[0, \frac{\pi}{2}]$. Speculators can discover the types of predictors in period 0 through a sequential search process. Each search round corresponds to a new exploration ("mining") of available data to obtain a new predictor. Each exploration costs c and returns a predictor with type θ drawn from $\phi(.)$ on $[\underline{\theta}, \frac{\pi}{2}]$ with probability $\alpha \leq 1$.[9] With probability $\alpha\text{Prob}(\theta < \underline{\theta}) + (1 - \alpha)$, the exploration

[8] Observe that the predictor s_θ is equivalent (in terms of informativeness) to the predictor $\hat{s}_\theta = \omega + (\cot(\theta))^{-1}\varepsilon_\theta$, whose precision is τ_θ/σ^2. Thus, a predictor of high quality is a predictor with high precision.

[9] For concreteness, one can interpret one exploration as running regressions of the asset payoff (e.g., stock earnings) on various variables (or combinations thereof) from a particular data set (e.g., financial statements or satellite images of retailers' parking lots). The theoretical R2 of a regression of ω on s_θ (i.e., $1 - \frac{Var(\omega|s_\theta)}{Var(\omega)}$) is equal to $cos(\theta)$. Thus, the higher the quality of a predictor, the higher the $R2$ of a regression of the asset payoff on the predictor. In other words, searching for predictors of high quality in the model is the same thing as searching for predictors with high R2s.

is unsuccessful. After each exploration, a speculator can decide to (i) stop searching and trade in period 1 on the predictor she just found (in which case, she learns the realization, s_θ, of her predictor in period 1) or (ii) start a new exploration in the hope of finding an even better predictor. We assume that there is no limit on the number of explorations.

Parameter $\underline{\theta}$ is the type of the best possible predictor that can be found given existing data. We refer to this parameter as being the *data frontier*. As explained in the introduction, data abundance should allow speculators to build even more informative predictors. This effect is captured in the model by a push back of the data frontier, i.e., a decrease in $\underline{\theta}$. For instance, the possibility for speculators to use satellite images of parking lots at Walmart combined with more traditional sources of information (e.g., financial statements) should allow investors to find more precise predictors of future earnings for Walmart. Data abundance might also create a needle in the haystack problem. New datasets open the possibility of obtaining even better predictors than in the past but this requires exploring these datasets first, at the risk of finding that they are useless. This problem becomes more acute as the diversity of datasets increases.[10] This effect is captured in the model by considering the effect of a decrease in α. In sum, we shall analyze the effect of data abundance by considering (separately) the effect of a decrease in $\underline{\theta}$ and α on equilibrium outcomes. Finally, parameter c represents the cost of exploring a specific dataset to identify a predictor. Greater computing power reduces this cost. For instance, with more powerful computers, one can explore more datasets in a fixed amount of time. So the time cost of data mining is smaller. Thus, we will analyze the effect of progress in computing power by considering the effect of a decrease in c on the equilibrium.

We focus on equilibria in which each speculator follows an optimal stopping rule θ_i^*. That is, speculator i stops searching for new predictors once she finds a predictor with type $\underline{\theta} \leq \theta < \theta_i^*$ (a predictor of sufficiently high quality in the feasible range). We denote by $\Lambda(\theta_i^*; \underline{\theta}, \alpha)$ the likelihood of this event (the probability of success) for speculator i in a given search round. That is:

$$\Lambda(\theta_i^*; \underline{\theta}, \alpha) \equiv \alpha \times \text{Prob}(\theta \in [\underline{\theta}, \theta_i^*]) = \alpha \times \int_{\underline{\theta}}^{\theta_i^*} \phi(\theta) d\theta.$$

Thus, a decrease in $\underline{\theta}$ raises the likelihood of finding a predictor in a given exploration, holding α

[10] See for instance "The quant fund investing in humans not algorithms" (AlphaVille, Financial Times, December 6, 2017), reporting discussions with a manager from TwoSigma noting that: "Data are noise. Drawing a tradable signal from that noise, meanwhile, takes work, since the signal is continuously evolving [...] Crucially, Duncombe added, there's qualitative data decay going on too. Back in the day, star managers may have had access to far smaller data sets, but the data in hand was of much higher quality."

constant. This effect captures the idea that while data abundance might reduce the fraction of informative datasets, it increases the chance of finding a good predictor once one has identified an informative dataset.

As the outcome of each exploration is random, the realized number of explorations varies across speculators. We denote by n_i the realized number of search rounds for speculator i. This number follows a geometric distribution with parameter $\Lambda(\theta_i^*, \underline{\theta})$. Thus, the expected number of explorations for a given speculator (a measure of her search intensity) is

$$\mathbb{E}[n_i] = \Lambda(\theta^*; \underline{\theta}, \alpha)^{-1}. \tag{2}$$

To simplify the exposition, we assume that speculators cannot "store" predictors that they turn down (i.e., the search for predictors is without recall). We show in the online appendix that this assumption is innocuous: The equilibrium of the model is identical if speculators have the option to pick the best predictor obtained up to the point at which they decide to stop searching for a predictor. Intuitively, the reason is that they face a stationary search problem because there is no limit on the number of explorations. Hence, if a speculator decides to keep searching after obtaining a given predictor in a given round, she must still find optimal to do so in each subsequent round (including for the best predictor obtained until this round). Second, we assume that speculators cannot trade on a combination of multiple predictors in period 1. One interpretation is that, if they wish to do so, they should first explore the quality of this combination at date 0 (that is, one can interpret s_θ as being a combination of multiple variables whose quality is higher than the quality of each variable taken independently).[11]

Trading phase. The trading phase starts after *all* speculators successfully complete their explorations (i.e., find a predictor with satisficing quality). At the beginning of period 1, each speculator observes the realization of her predictor, s_θ and choose a trading strategy, i.e., demand schedule, $x_i(s_\theta, p)$, where, p, is the asset price in period 1.

We formalize trading as in Vives (1995). That is, we assume that trading takes place between the speculators, noise traders and a risk-neutral market maker. Noise traders' aggregate demand is price-inelastic and equal to η, where $\eta \sim \mathcal{N}(0, \nu^2)$ (and η is independent of ω and errors' in speculators' signals). The market-maker behaves competitively. He observes investors'

[11] For instance, using a combination of variables based on parking-lots satellite images and credit-card receipts (wto distinct datasets) might generate a better predictor of retailers' earnings than using variables independently.

aggregate demand, $D(p) = \int x_i(s_\theta, p)di + \eta$ and sets a price equal to his expectation of the asset payoff (so that his expected profit is zero):

$$p = \mathbb{E}\left[\omega \,|D(p)\right]. \tag{3}$$

Speculators' objective function. At $t = 2$, the asset pays off and speculator i's final wealth is

$$W_i = x_i(s_\theta, p)(\omega - p) - n_i c. \tag{4}$$

The number of explorations for speculator i, n_i, is independent from the asset payoff, its price, and the realization of the speculator's predictor s_θ since n_i is determined in period 0, before the realizations of these variables. Thus, the ex-ante expected utility of a speculator can be written:

$$\mathbb{E}\left[-\exp(-\rho W_i)\right] = \underbrace{\mathbb{E}\left[-\exp(-\rho(x_i(s_\theta, p)(\omega - p))\right]}_{\text{Expected Utility from Trading}} \times \underbrace{\mathbb{E}\left[\exp(\rho(n_i c))\right]}_{\text{Expected Utility Cost of Exploration}} \tag{5}$$

The first term in this expression represents the ex-ante expected utility that a speculator derives from trading while the second term represents the expected utility of the cost of explorations. The first term depends both on the investor's optimal trading strategy ($x_i(s_{\theta,i}, p)$) and her optimal stopping rule (θ_i^*) because this rule determines the distribution of s_{theta}. The second term depends on the speculator's stopping rule, θ_i^* because it determines the distribution of n_i. Each speculator chooses her stopping rule, θ_i^*, and her trading strategy, $x_i(s_{\theta,i}, p)$, to maximize her ex-ante expected utility. This optimization problem is dynamic because the speculator's trading strategy can be contingent on the realization of the speculator's predictor and the asset price in period 1 while the stopping rule must be chosen in period 0.

3. Data Abundance, Computing Power, and Optimal Data Mining

3.1 Equilibrium

We focus on symmetric equilibria in which all speculators choose the same stopping rule, θ^*. We solve for such an equilibrium as follows. First, given the equilibrium outcome of the exploration phase in period 0, we first solve for the equilibrium of the trading phase in period 1. Armed with this result, we compute the expected utility achieved by a speculator i who chooses a predictor of type θ in period 0 (before she observes the realization of her predictor) and we deduce the

optimal stopping rule of the speculator $\theta_i^*(\theta^*)$, when she expects other speculators to follow the stopping rule θ^*. Finally, we pin down θ^* by observing that, in a symmetric equilibrium, the speculator's best response must be identical to other speculators' stopping rule, i.e., $\theta_i^*(\theta^*) = \theta^*$.

Equilibrium of the asset market in period 1. The outcome of the exploration phase is characterized by the distribution of the θ's for the predictors obtained by speculators. Let $\phi^*(\theta; \theta^*, \underline{\theta})$ be this distribution given that speculators' follow the stopping rule θ^*:

$$\phi^*(\theta; \theta_i^*; \underline{\theta}, \alpha) = \frac{\phi(\theta)}{\Lambda(\theta_i^*; \underline{\theta}, \alpha)}. \tag{6}$$

This distribution characterizes the heterogeneity of speculators' predictors in equilibrium. We denote the *average* quality of predictors across all speculators in period 1 by $\bar{\tau}(\theta^*, \underline{\theta}) = \mathbb{E}\left[\tau_\theta | \, \underline{\theta} \leq \theta \leq \theta^*\right]$. We assume that the distribution of predictors' types is such that $\bar{\tau}(\theta^*; 0, \alpha)$ exists.[12] Proposition 1 provides the equilibrium of the asset market in period 1.

Proposition 1. *In period* 1*, the equilibrium trading strategy of a speculator with type* θ *is:*

$$x^*(s_\theta, p) = \frac{\mathbb{E}[\omega | s_\theta, p] - p}{\rho \mathbb{V}[\omega | s_\theta, p]} = \frac{\tau(\theta)}{\rho \sigma^2} \left(\hat{s}_\theta - p\right), \tag{7}$$

where $\hat{s}_\theta = \omega + \cot^{-1}(\theta)\varepsilon_\theta$ *and the equilibrium price of the asset is:*

$$p^* = \mathbb{E}[\omega | D(p)] = \lambda(\theta^*)\xi. \tag{8}$$

where

$$\xi = \omega + \rho\sigma^2 \bar{\tau}^{-1}(\theta^*; \underline{\theta}, \alpha)\eta, \quad \textit{and} \quad \lambda(\theta^*) = \frac{\bar{\tau}(\theta^*; \underline{\theta}, \alpha)^2}{\bar{\tau}(\theta^*; \underline{\theta}, \alpha)^2 + \rho^2\sigma^2\nu^2}, \tag{9}$$

This result extends Proposition 1.1 in Vives (1995) to the case in which speculators have signals of heterogenous precisions (determined by their θ in our model). The predictor s_θ is informationally equivalent to the predictor $\hat{s}_\theta = \omega + \cot^{-1}(\theta)\varepsilon_\theta$. A speculator's optimal position in the asset is equal to the difference between $\hat{s}_\theta$ and the price of the asset (her expected dollar return) scaled by a factor that increases with the quality of the predictor and decreases with the speculator's risk aversion. The scaling factor measures the speculator's aggressiveness in trading on her predictor. Speculators with predictors of higher quality trade more aggressively

[12] The reason for this technical condition is that we want to consider what happens when $\underline{\theta}$ goes to zero. In this case, $\bar{\tau}(\theta^*, \underline{\theta})$ is not necessarily defined for all $\phi(.)$ because the quality of a predictor becomes infinite as its type goes to zero.

on their signal because they face less risk (their forecast of the asset payoff is more accurate).

The total demand for the asset ($D(p)$) aggregates speculators' orders and therefore reflects their information. Observing this demand is informationally equivalent to observing the signal ξ, whose informativeness increases with the average quality of speculators' predictors, $\bar{\tau}(\theta^*; \underline{\theta}, \alpha)$. Thus, the market maker can form a more accurate forecast of the asset payoff and the asset price is therefore be more informative when the average quality of speculators' predictors, $\bar{\tau}(\theta^*, \underline{\theta})$, is higher. To see this, let measure the informativeness of the asset price by $\mathcal{I}(\theta^*; \underline{\theta}, \alpha) = \mathbb{V}[\omega \mid p^*]^{-1}$ as in Grossman and Stiglitz (1980). Using Proposition 1, we obtain:

$$\mathcal{I}(\theta^*; \underline{\theta}, \alpha) = \tau_\omega + \frac{\bar{\tau}(\theta^*, \underline{\theta}, \alpha)^2 \tau_\omega^2}{\rho^2 \nu^2}, \tag{10}$$

where $\tau_\omega = \frac{1}{\sigma^2}$ is the precision of speculators' prior about the asset payoff. As already explained, the asset price is more informative when the average quality of speculator's predictors increases. Thus, the informativeness of the asset price is inversely related to θ^* because $\bar{\tau}(\theta^*; \underline{\theta}, \alpha)$ decreases with θ^* (remember that the quality of a predictor is inversely related to θ). Thus, other things equal, price informativeness is smaller when speculators chooses a less stringent stopping rule for the quality of the predictors on which they trade.

Equilibrium of the exploration phase. Using the characterization of the equilibrium of the asset market and the fact that $\mathbb{E}[x(s_\theta, p)] = 0$ (Proposition 1), we can compute a speculator's expected utility from trading ex-ante, i.e., before observing the realization of her predictor and and the equilibrium price, when her predictor has type θ and other speculators follow the stopping rule θ^*. We denote this ex-ante expected utility by $g(\theta, \theta^*)$ and refer to it as the trading value of a predictor with type θ for brevity. Formally:

$$g(\theta, \theta^*) \equiv \mathbb{E}\left[-\exp(-\rho(x^*(s_\theta, p^*)(\omega - p^*)) \mid \theta_i = \theta\right]. \tag{11}$$

Lemma 1. *In equilibrium, the trading value of a predictor with type θ is:*

$$g(\theta, \theta^*) = -\left(1 + \rho^2 \left(\frac{\mathbb{V}[\mathbb{E}(\omega | s_\theta, p) - p]}{\mathbb{V}[\omega | s_\theta, p]}\right)\right)^{\left(-\frac{1}{2}\right)} = -\left(1 + \frac{\tau(\theta)\tau_\omega}{\mathcal{I}(\theta^*; \underline{\theta}, \alpha)}\right)^{-\frac{1}{2}}. \tag{12}$$

Thus, the trading value of a predictor increases (is closer to zero) with its quality and decreases with the informativeness of the asset price. Thus, the trading value of a predictor is

inversely related to the average quality of predictors used by speculators. Hence, the value of a given predictor for a speculator depends on the search strategy followed by other speculators: It is smaller if other speculators are more demanding for the quality of their predictors (i.e., when θ^* decreases).

Armed with Lemma 1, we can now derive a speculator's optimal stopping rule given that other speculators follow the stopping rule θ^*. Let $\hat{\theta}_i$ be an arbitrary stopping rule for speculator i. The speculator's continuation utility (the expected utility of launching a new round of exploration) after turning down a predictor is:

$$J(\widehat{\theta}_i, \theta^*) = \exp(\rho c) \left(\Lambda(\widehat{\theta}_i; \underline{\theta}, \alpha) \mathbb{E} \left[g(\theta, \theta^*) | \, \underline{\theta} \leq \theta \leq \widehat{\theta}_i \right] + (1 - \Lambda(\widehat{\theta}_i; \underline{\theta}, \alpha)) J(\widehat{\theta}_i, \theta^*) \right) \tag{13}$$

The first term ($\exp(\rho c)$) in eq.(13) is the expected utility cost of running an additional search . The second term is the likelihood that the next exploration is successful times the average trading value of a predictor conditional on the type θ of this predictor being satisficing (i.e., in $[\underline{\theta}, \widehat{\theta}_i]$). Finally, the third term is the likelihood that the next exploration is unsuccessful time the speculator's continuation utility when she turns down a predictor. Solving eq.(13) for $J(\widehat{\theta}_i, \theta^*)$, we obtain:

$$J(\widehat{\theta}_i, \theta^*) = \underbrace{\left[\frac{\exp(\rho c) \Lambda(\widehat{\theta}_i; \underline{\theta}, \alpha)}{1 - \exp(\rho c)(1 - \Lambda(\widehat{\theta}_i; \underline{\theta}, \alpha))} \right]}_{\text{Expected Utility Cost from Exploration}} \times \underbrace{\mathbb{E} \left[g(\theta, \theta^*) | \, \underline{\theta} \leq \theta \leq \widehat{\theta}_i \right]}_{\text{Expected Utility from Trading}} \tag{14}$$

The continuation value of the speculator when she turns down a predictor does not depend on the outcomes of past explorations because these outcomes do not affect the speculator's opportunity set in future explorations. Thus, $J(\widehat{\theta}_i, \theta^*)$ is also the speculator's ex-ante expected utility before starting any exploration in period 0. As explained previously, it is the product of the expected utility cost from explorations and the expected utility from trading.

Now suppose that speculator i has obtained a predictor with quality θ. If the speculator stops exploring the data at this stage, her expected utility is $g(\theta, \theta^*)$ (her cost of exploration to obtain this predictor is sunk). If instead the speculator decides to launch a new round of exploration, her expected utility is $J(\widehat{\theta}_i, \theta^*)$. Thus, the optimal decision is to stop searching for a predictor if $g(\theta, \theta^*) \geq J(\widehat{\theta}_i, \theta^*)$ and to keep searching otherwise. As $g(\theta, \theta^*)$ decreases with θ, the optimal stopping rule of the speculator, θ_i^*, is the value of θ such that the speculator is just

indifferent between these two options. Thus, θ_i^* is the value of $\widehat{\theta}_i$ that solves:

$$g(\theta_i^*, \theta^*) = J(\theta_i^*, \theta^*). \tag{15}$$

We show in the proof of Proposition 2 that the solution to this equation, $\theta_i^*(\theta^*) = \theta^*$, is unique. In a symmetric equilibrium, it must be that $\theta_i^*(\theta^*) = \theta^*$. We deduce that θ^* solves:

$$g(\theta^*, \theta^*) = J(\theta^*, \theta^*). \tag{16}$$

Using the expression for $J(., \theta^*)$ in eq.(13), we can equivalently rewrite this equilibrium condition as:

$$F(\theta^*) = \exp(-\rho c), \tag{17}$$

where $F(.)$ is defined as:

$$F(\theta^*) = \int_{\underline{\theta}}^{\theta^*} r(\theta, \theta^*)\phi(\theta)d\theta + (1 - \Lambda(\theta^*; \underline{\theta}, \alpha)), \quad \text{for } \theta^* \in [\underline{\theta}, \pi/2], \tag{18}$$

with

$$r(\theta, \theta^*) \equiv \frac{g(\theta, \theta^*)}{g(\theta^*, \theta^*)} = \left(\frac{\tau(\theta^*)\tau_\omega + \mathcal{I}(\theta^*; \underline{\theta}, \alpha)}{\tau(\theta)\tau_\omega + \mathcal{I}(\theta^*; \underline{\theta}, \alpha)}\right)^{\frac{1}{2}}, \tag{19}$$

where the second equality follows from eq.(12). Thus, $r(\theta, \theta^*)$ is the ratio of the trading value of a predictor with type $\theta \in [\underline{\theta}, \theta^*]$ to the trading value of a predictor with type θ^*. It is less than 1 because the trading value of a predictor increases with its quality ($g(\theta^*, \theta^*) < g(\theta, \theta^*) < 0$).

We show in the proof of Proposition 2 that $F(\theta^*)$ decreases with θ^* because the ratio $r(\theta, \theta^*)$ decreases with θ^*. Indeed, an increase in θ^* (i) reduces the trading value of a predictor of type θ^* (the denominator of $r(\theta, \theta^*)$) since its quality declines and (ii) increases the trading value of a predictor with type $\theta < \theta^*$ (the numerator of $r(\theta, \theta^*)$) because price informativeness decreases with θ^*. We have $F(\underline{\theta}) = 1$, $0 < F(\pi/2) < 1$ and $\exp(-\rho c) < 1$ (since $c > 0$). Thus, there are two possibilities. If $\exp(-\rho c) \geq F(\pi/2)$, there is a unique solution to eq.(17) and this solution is in $(\underline{\theta}, \pi/2)$. If instead $\exp(-\rho c) \leq F(\pi/2)$ then there is no solution. In this case, the cost of exploration is so large that there is no symmetric equilibrium in which all speculators are active (i.e., search for a predictor in period 0). However, in this case, one can build an equilibrium in which only a fraction of all speculators are active. Active speculators search for a predictor with a stopping rule equal to $\theta^* = \pi/2$ while others remain completely inactive. In

this equilibrium all speculators are indifferent between being active or not. The case in which all speculators are active is more interesting. Thus, henceforth we assume that c is low enough so that $\exp(-\rho c) \geq F(\pi/2)$. These observations yield the following result.

Proposition 2. *There is a unique symmetric equilibrium of the exploration phase in which all speculators are active (i.e., a unique stopping rule $\theta^* < \frac{\pi}{2}$ common to all speculators) if and only if* $\exp(-\rho c) \geq F(\pi/2)$.

3.2 How do data abundance and computing power affect speculators' search for predictors?

In this section, using the previous results, we analyze how data abundance (a decrease in $\underline{\theta}$ and/or α) and computing power (a decrease in c) affect speculators' optimal stopping rule (θ^*) in equilibrium. For brevity, we say that speculators optimally choose to trade on *less* (more) demanding predictors when θ^* *increases* (decreases). The next proposition shows that greater computing power (a drop in c) always leads speculators to be more demanding for the quality of their predictors.

Proposition 3. *A decrease in the cost of exploration, c, always reduces the stopping rule θ^* used by speculators in equilibrium ($\frac{\partial \theta^*}{\partial c} > 0$). Thus, greater computing power raises the quality, $\tau(\theta^*)$, of the worst predictor used by speculators in equilibrium.*

The economic mechanism for this finding is as follows. Holding θ^* constant, a decrease in the per-exploration cost, c, directly reduces the expected utility cost of launching a new exploration after finding a predictor (the first term in bracket in eq.(14)). Hence, it raises the value of searching for another predictor after finding one (i.e., $J(\theta^*, \theta^*)$). This direct effect induces speculators to be more demanding for the quality of their predictor and therefore works to decrease θ^*. One indirect effect of this behavior is that, on average, speculators trade more aggressively on their signal (the "competition effect") because they face less uncertainty on the asset payoff since their predictors are better on average. As a result, price informativeness increases. This indirect effect of a decrease in the per exploration cost reduces the expected utility from trading on a satisficing predictor (the second term in bracket in eq.(14)) and therefore dampens the direct positive effect of a decrease in c on the value of searching for a better predictor after finding one. However, it is never strong enough to fully offset the decrease in the total expected search cost due to a decrease in c.

We now consider the effect of data abundance on speculators' optimal stopping rule. Remember that data abundance has two consequences in the model: (i) it pushes back the data frontier by raising the quality of the best predictor and (ii) it increases the risk for speculators of using datasets, which after exploration proves to be useless (the needle in the haystack problem).

Proposition 4.

1. *A decrease in the fraction of informative datasets, α, always increases speculators' stopping rule, θ^*, in equilibrium ($\frac{\partial \theta^*}{\partial \alpha} > 0$). Thus, the needle in the haystack problem reduces the quality, $\tau(\theta^*)$, of the worst predictor used by speculators in equilibrium.*

2. *The effect of a decrease in $\underline{\theta}$ on speculators' stopping rule is ambiguous. However, when $\underline{\theta}$ is less than $\underline{\theta}^{tr}(c)$, a decrease in $\underline{\theta}$ always increases speculators' stopping rule in equilibrium ($\frac{\partial \theta^*}{\partial \underline{\theta}} > 0$ for $\underline{\theta} < \underline{\theta}^{tr}(c)$) and reduces the quality, $\tau(\theta^*)$, of the worst predictor used by speculators in equilibrium.*

Thus, when the needle in the haystack problem becomes more acute, speculators become less demanding for the quality of their predictors. Intuitively, a drop in α increases the expected utility cost of launching a new exploration after finding a predictor (the first term in bracket in eq.(14)) because it reduces the likelihood of finding a predictor in a given exploration (Λ). Thus, after turning down a predictor, speculators expect to go through a larger number of explorations rounds before finding a satisficing predictor and therefore to pay a larger total cost of search. This direct effect induces speculators to be less demanding for the quality of their predictor and therefore works to increase θ^* (reduce $\tau(\theta^*)$). Indirectly, this behavior reduces asset price informativeness and therefore raises the expected utility from trading on a satisficing predictor (the second term in bracket in eq.(14)), which alleviates the direct negative effect of a decrease in α on the value of searching for a better predictor after finding one. However, this indirect effect is never strong enough to fully offset the increase in the total expected utility cost of search due to a decrease in α. In sum, qualitatively, the effect of a drop in α is qualitatively similar to that of an increase in the per exploration cost.

The effect of pushing back the data frontier on speculators' stopping rule is more complex. Counterintuitively, it can lead speculators to trade on predictors of worse quality, even though the quality of the best predictor increases. The reason is as follows. Remember that the continuation value of searching for a predictor is the product of the expected utility cost of

search times the expected utiliy from trading on a satisficing predictor (see eq.(14). On the one hand, pushing back the data frontier increases the chance of finding a satisficing predictor conditional on finding an informative dataset ($\Lambda(\theta^*;\underline{\theta},\alpha)$ increases when $\underline{\theta}$ goes down). This effect accelerates the speed at which speculators find a predictor and therefore reduces the expected utility cost of searching for a new predictor after rejecting one. This effect tends to increase the continuation value of searching for a predictor. On the other hand, a push back of the data frontier affects the expected utility from trading for two reasons. First, it gives the possibility to obtain more informative predictors than those existing before ("the hidden gold nugget effect"), which raises the expected utility from trading on a satisficing predictor. Second, it increases price informativeness (other things equal, $\mathcal{I}(\theta^*;\underline{\theta},\alpha)$ increases when $\underline{\theta}$ decreases) because speculators who obtain the new most informative predictors trade even more aggressively than all other speculators. As a result, speculators' aggregate demand and therefore the asset price are more informative, which reduces the value of being informed. This effect (that we call the competition effect of data abundance) reduces the expected utility from trading on a satisficing predictor. Thus, the sign of a change in the data frontier on the expected utility from trading is ambiguous.

To see this more formally, observe that:

$$\frac{\partial \mathbb{E}\left[g(\theta,\theta^*)|\,\underline{\theta}\leq\theta\leq\theta^*\right]}{\partial\underline{\theta}}=\frac{\phi(\underline{\theta})}{\Lambda}\left[\underbrace{\left[g(\theta,\theta^*)|\,\underline{\theta}\leq\theta\leq\widehat{\theta}_i\right]-(g(\underline{\theta},\theta^*)}_{\text{Hidden Gold Nugget Effect; }<0}+\underbrace{\int_{\underline{\theta}}^{\theta^*}\frac{\partial g(\theta,\theta^*)}{\partial\underline{\theta}}\phi(\theta)d\theta}_{\text{Competition Effect;}>0}\right] \tag{20}$$

When $\underline{\theta}$ becomes small enough, the competition effect dominates the hidden gold nugget effect and the expected utility from trading on a satisficing predictor drops. The second part of Proposition 4 shows that there is always a sufficiently low value of $\underline{\theta}$ such that this drop more offsets the reduction in the expected utiulity cost of finding a predictor. When this happens, pushing back the frontier further reduces the continuation value of exploration. Hence, speculator choose a less stringent stopping rule in equilibrium and some optimally choose to trade on less informative predictors ($\tau(\theta^*)$ decreases).

We illustrate Proposition 4 by considering two particular specifications of the density of θ. In specification 1, we assume that $\phi(\theta)=3\cos(\theta)\sin^2(\theta)$ while in specification 2 we assume that $\phi(\theta)=5cos(\theta)sin^4(\theta)$. These specifications are convenient because they enable us to compute all variables of interest in closed forms (see the internet appendix). The main difference between

specifications 1 and 2 is that the distribution of θ has a much fatter right-tail in the first case (see the internet appendix). Figure 1 below shows the effect of a change in the exploration cost (c) and the data frontier ($\underline{\theta}$) on the equilibrium value of θ^*. In either case, as implied by Proposition 4, a push back of the data frontier initially raises the quality of the worst predictor used by speculators in equilibrium (reduces θ^*) but, eventually, at some point this effect is reversed. This reversal is more pronounced in specification 1 than in specification 2, which shows that the magnitude of the effect depends on the exact distribution of predictors' quality.

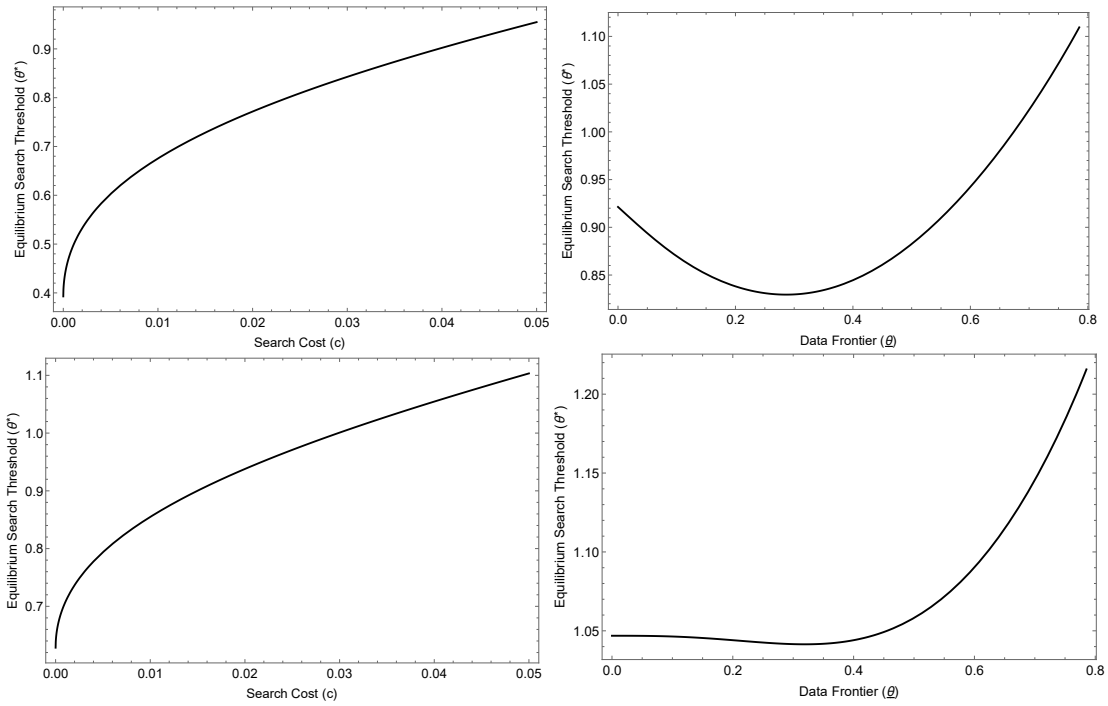

Figure 1: In each figure, the left hand-side graph represents the equilibrium search threshold, θ^*, as a function of the search cost, c, with other parameter values, $\underline{\theta} = \pi/8, \rho = 1, \sigma^2 = 1, \nu^2 = 1$. The right hand-side graph represents the equilibrium search threshold, θ^*, as a function of the data frontier, $\underline{\theta}$, with other parameter values, $c = 0.03, \rho = 1, \sigma^2 = 1, \nu^2 = 1$. In the first figure, we assume that $\phi(\theta) = 3cos(\theta)sin^2(\theta)$ (Case 1) while in the second we assume that $\phi(\theta) = 5cos(\theta)sin^4(\theta)$ (Case 2).

Proposition 5. *In equilibrium, the quality of the worst predictor used in equilibrium, $\tau(\theta^*)$, increases with the volume of noise trading, ν^2, or the volatility of the asset payoff, σ^2.*

An increase in the volume of noise trading (measured by ν^2) reduces the informativeness of the equilibrium price. Other things equal, this means that the trading value of each predictor relative to the worst predictor used in equilibrium gets larger (that is $r(\theta, \theta^*)$ goes up). This effect induces all speculators to search more intensively, i.e., raises the bar for the minimal quality of the predictors on which speculators trade. The intuition for the effect of the volatility

of the asset is identical. Thus, the model implies that the average precision of predictors and the intensity for the search of predictors should be larger in assets that attract more attention from noise traders or more volatile assets. Simultaneously, the diversity of predictors for these assets should be smaller.

4. Testable Implications

4.1 Data Abundance, Computing Power, and Stock Picking Abilities

In order to test implications derived in the previous section, empiricists need to estimate the quality of predictors used by speculators. One way to do so is to regress the position of each speculator i (or their net order flow) in a given asset, $x_i(s_\theta, p^*)$, on the asset return $(\omega - p^*)$ (e.g., using quarterly data on mutual funds' positions). Indeed, the coefficient of this regression, β_θ is given by:

$$\beta_\theta = \frac{Cov(x(s_\theta), p^*), \omega - p^*)}{Var(\omega - p^*)} = \frac{\tau(\theta)}{\rho}, \tag{21}$$

where the last equality follows from Proposition 1. Intuitively, β_θ is a measure of a speculator's stock picking ability.[13] Equation (21) shows that, holding risk aversion constant, a ranking of speculators based on their stock picking ability (measured by β_θ) is identical to a ranking based on the (unobservable) quality of their predictors, $\tau(\theta)$. This is intuitive: speculators with better predictors should display a better stock picking ability. Moreover, the cross-sectional dispersion in βs' is proportional to the cross-sectional dispersion in the quality of speculators' predictors. Thus, one could test the implications derived in this section by ranking speculators (e.g., quantitative asset managers) based on their stock picking ability (measured by βs) and test whether shocks to computing power or data abundance have the effects predicted by Propositions 3 and 4. For instance, one could test whether positive shocks to computing power increase the stock picking ability (measured by β) of the funds with the lowest βs' (say in the lowest decile) or whether the introduction of new datasets that increases funds' ability to build more informative predictors (e.g., the creation of firms selling satellite images to investors, as in Zhu (2019)) have a *negative* impact on this ability (while improving the stock picking ability of the best performing funds). One could also test whether periods of high volatility are associated with a decrease in the difference between the stock ability of speculators in lowest decile of β

[13] Kacperczyk et al. (2016) measure mutual funds' stock picing ability in a similar way. See Section 2.1 in their paper.

(those with predictors of low quality) and the top decile of β (those with the most informative predictors), as implied by Proposition 5.

4.2 Data Abundance, Computing Power, and Asset Price Informativeness

There is a debate about the effects of progress in information technologies on asset price informativeness. For instance Bai et al. (2016) find that the price SP500 stocks has become more informative since 1960 and attribute this evolution to improvements in information technologies. Farboodi et al. (2019) find the opposite pattern for all U.S. stocks, except for large growth stocks. They argue that this evolution reflects the fact that investors have optimally increased their production of information about large, fast growing firms at the expense of other firms.

Information technologies improve investors' ability to build more accurate predictors of asset payoffs in two ways. On the one hand, they reduce the cost of filtering out noise from raw data (e.g., greater computing power is key for implementing techniques of artificial intelligence such as deep neural networks). On the other, they offer the possibility to generate more diverse data (e.g., through image capture techniques) and to store these data more efficiently. Our model suggests that these two different aspects of information technologies do not affect asset price informativeness in the same way. We first discuss the effect of greater computing power.

Proposition 6. *In equilibrium, an increase in computing power (a decrease in c) raises the average quality of speculators' predictors and therefore price informativeness.*

As explained previously, greater computing power induces speculators to be more demanding for the quality of their predictors (to put more effort in the search of good predictors) because it reduces the cost of exploring new data to obtain a predictor. As a result, speculators obtain signals of higher quality on average and therefore trade more aggressively on their signals on average. In turn, their aggregate demand for an asset is more informative and therefore price informativeness increases.

Proposition 7.

1. *In equilibrium, an improvement in the quality of the most informative predictor (a decrease in $\underline{\theta}$) raises the average quality of speculators' predictors and therefore price informativeness.*

2. *In equilibrium, a decrease in the proportion of informative datasets (a decrease in α) reduces the average quality of speculators' predictors and therefore price informativeness.*

Thus the effect of data abundance on price informativeness is ambiguous. If data abundance only pushes back the data frontier, i.e., improves the quality of the most informative predictor, then it always improves asset price informativeness, even when it induces speculators to be less demanding for the quality of their predictors (i.e., when a decrease in $\underline{\theta}$ reduces $\tau(\theta^*)$). The reason is that the drop in the quality of the worst predictor used in equilibrium (if it happens) is always smaller than the improvement in the quality of the best predictor in equilibrium. As a result, a push back of the data frontier raises the average quality of predictors and the average trading aggressiveness of speculators. In contrast, if data abundance only makes it more difficult to identify truly informative datasets, it unambiguously leads speculators to be less demanding for the quality of their predictors without changing the quality of the best predictor. As a result the average quality of predictors drops, speculators' aggregate demand is less informative and therefore price informativeness drops.

Of course, in reality, these two effects of data abundance operate jointly. As a result, the net effect of data abundance on the long run evolution of asset price informativeness is ambiguous, as found empirically by Bai et al. (2016) and Farboodi et al. (2019). Figure 2 illustrates this point with a numerical example. In this example we assume that α increases with $\underline{\theta}$ (specifically, we specify α as $\alpha = Min\{1, 0.32 + 0.8 \times \underline{\theta}\}$). Thus, data abundance (a drop in $\underline{\theta}$) generates both an increase in the quality of the best predictor and a needle in the haystack problem (a drop in the fraction of informative datasets). As shown by Figure 2, when these two dimensions of data abundance operate jointly, price informativeness initially rises with data abundance (starting from a large $\underline{\theta}$) until it reaches a peak after which it decreases. The reason is that when $\underline{\theta}$, both dimensions of data abundance induce speculators to be less demanding for the quality of their predictors, which eventually is harmful for price informativeness.

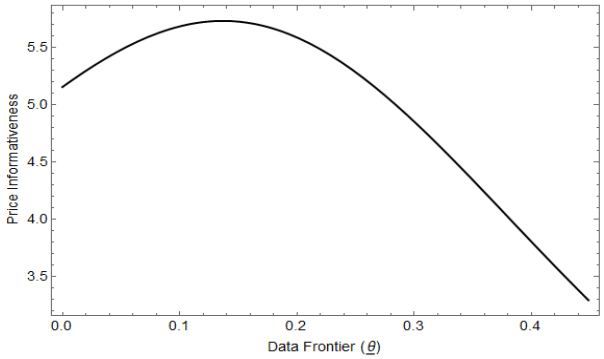

Figure 2: This graph shows the evolution of price informativeness in equilibrium, $\mathcal{I}(\theta^*, \underline{\theta})$ as a function of the data frontier, $\underline{\theta}$ when $\phi(\theta) = 3cos(\theta)sin^2(\theta)$ and $\alpha = Min1, 0.32 + 0.8 * \underline{\theta}$. Other parameter values, $c = 0.03, \rho = 1, \sigma^2 = 1, \nu^2 = 1$.

4.3 Data abundance, Computing Power and Trading Profits

In this section, we analyze how data abundance and computing power affects (i) the distribution of trading profits for speculators and (ii) the correlation in their trades. In equilibrium, the total trading profit ("excess return"), $\pi(s_\theta)$, of a speculator with type θ on his position in the risky asset is:

$$\pi(s_\theta) = x^*(s_\theta, p^*) \times (\omega - p^*), \tag{22}$$

where $x^*(s_\theta, p^*)$ and p^* are given by eq.(7) and eq.(8), respectively. Using eq.(7), we deduce that:

$$x^*(s_\theta, p^*) = \frac{1}{\rho\sigma^2} \left(\tau(\theta)(\omega - p^*) + \cot(\theta)\varepsilon_\theta\right). \tag{23}$$

Thus, the *expected* trading profit of a speculator with type θ is:

$$\pi(\theta) = \mathbb{E}[\pi(s_\theta)|\theta] = \frac{\tau(\theta)}{\rho\sigma^2}\mathbb{V}[\omega - p^* \mid \theta] = \frac{\tau(\theta)}{\rho\sigma^2\mathcal{I}(\theta^*, \underline{\theta})} \tag{24}$$

It follows that the unconditional expected trading profit for speculators is:

$$\mathbb{E}[\pi(\theta)] = \frac{\bar{\tau}(\theta^*, \underline{\theta})}{\rho\sigma^2\mathcal{I}(\theta^*, \underline{\theta})} = \frac{1}{\rho\sigma^2}\left(\frac{\tau_\omega}{\bar{\tau}(\theta^*, \underline{\theta})} + \frac{\bar{\tau}(\theta^*, \underline{\theta})}{\rho^2\nu^2}\right)^{-1}, \tag{25}$$

and the variance of trading profits for speculators is:

$$\mathbb{V}[\pi(\theta)] = \frac{\mathbb{V}[\tau(\theta) \mid \underline{\theta} < \theta < \theta^*]}{\sigma^4 \rho^2 \mathcal{I}^2(\theta^*, \underline{\theta})}. \tag{26}$$

Empirically, $\mathbb{E}[\pi(\theta)]$ and $\mathbb{V}[\pi(\theta)]$ could be measured by the cross-sectional mean and variance of the trading profits of quantitative funds (over, for instance, a given quarter).

Observe that an increase in the average quality of predictors ($\bar{\tau}(\theta^*, \underline{\theta})$ has an ambiguous effect on the expected profit of speculators. One the one hand, this increase improves speculators's stock picking ability (the direction of their position is more correlated with the subsequent asset return). On the other hand, it increases asset price informativeness because it makes speculators' aggregate demand more informative (as speculators trade more aggressively on their signals on average when these are more accurate). As shown by eq.(25), the first effect raises speculators' expected profit while the second reduces it. Using eq.(25), it is easily shown that the first effect dominates if and only if $\bar{\tau}(\theta^*, \underline{\theta}) \leq \tau_\omega \rho^2 \nu^2$. We deduce the following result.

Proposition 8. *Suppose $\underline{\theta} > \tau_\omega \rho^2 \nu^2$.*

1. *Let $\hat{c}$ be the unique solution to $\bar{\tau}(\theta^*(\underline{\theta}, \hat{c}, \alpha), \underline{\theta}) = \tau_\omega \rho^2 \nu^2$. Holding all parameters constant, except c, speculators expected profit reaches its maximum for $c = \hat{c}$.*

2. *Let $\hat{\underline{\theta}}$ be the unique solution to $\bar{\tau}(\theta^*(\hat{\underline{\theta}}, c, \alpha), \underline{\theta}) = \tau_\omega \rho^2 \nu^2$. Holding all parameters constant, except $\underline{\theta}$, speculators' expected profit reaches its maximum for $\underline{\theta} = \hat{\underline{\theta}}$.*

3. *Let $\hat{\alpha}$ be the unique solution to $\bar{\tau}(\theta^*(\underline{\theta}, c, \hat{\alpha}), \underline{\theta}) = \tau_\omega \rho^2 \nu^2$. Holding all parameters constant, except α, speculators' expected profit reaches its maximum for $\alpha = \hat{\alpha}$.*

If $\underline{\theta} \leq \tau_\omega \rho^2 \nu^2$, more data or greater computing power always increases speculators' average profit.

Thus, when $\underline{\theta} > \tau_\omega \rho^2 \nu^2$, an increase in computing power and data abundance has an ambiguous effect on speculators' expected profits because a decrease in c, $\underline{\theta}$, or α affect (i) the average quality of speculators' signals and (ii) price informativeness in the same direction (e.g., both a decrease in c or $\underline{\theta}$ raise both the average qualityt of speculators' signals and price informativeness). However, when computing power is high enough ($c < \hat{c}$) or data are sufficiently abundant ($\alpha < \hat{\alpha}$ or $\underline{\theta} < \hat{\underline{\theta}}$, greater computing power or data abundance have a negative effect on speculators' expected profit, either because it makes price too informative ($\underline{\theta} < \hat{\underline{\theta}}$ or $c < \hat{c}$)

or because the needle in the haystack problem reduces too much speculators' incentive to search for good predictors, so that their average signal are of too poor quality ($\alpha < \hat{\alpha}$).

Now consider the effect of changes in the cost of processing data and data abundance on the dispersion ($\mathbb{V}[\pi(\theta)]$) of expected trading profits across speculators. Using eq.(26), we obtain the following result.

Proposition 9.

1. *Other things equal, the dispersion of speculators' expected trading profit increases when the cost of processing data goes down for c small enough ($\frac{\partial \mathbb{V}[\pi(\theta)]}{\partial c} > 0$ for c sufficiently close to zero).*

2. *Other things equal, the dispersion of speculators' expected profit decreases when the data frontier is pushed back for $\underline{\theta}$ small enough ($\frac{\partial \mathbb{V}[\pi(\theta)]}{\partial \underline{\theta}} < 0$ for $\underline{\theta}$ sufficiently close to zero).*

3. *Other things equal, the dispersion of speculators' expected trading profit increases when the fraction of informative datasets decreases (α decreases).*

To understand the first part of the proposition, suppose that $c = 0$. In this case, all speculators search for a predictor until they find one with the highest possible quality, $\theta^* = \underline{\theta}$. As a result, all speculators trade on predictors of the same quality ($\mathbb{V}[\tau(\theta) \mid \underline{\theta} < \theta < \theta^*] = 0$) and therefore the dispersion of expected trading profits is nil, as can be seen by inspection of the expression for $\mathbb{V}[\pi(\theta)]$ (eq.(26)). Now consider a small increase in c starting from the situation in which $c = 0$. This increase raises θ^* and therefore the dispersion of the quality of predictors used by speculators ($\mathbb{V}[\tau(\theta) \mid \underline{\theta} < \theta < \theta^*]$ increases). As a result, the dispersion of trading profits increases as well. This increase is amplified by the fact that price informativeness goes down, which works to increase the dispersion in trading profits as well (see the expression for $\mathbb{V}[\pi(\theta)]$ in eq.(26)). As these effects still hold for larger values of c, we conjecture that the first part of Proposition 9 holds for all values of c but we have not been able to show it analytically (numerical simulations suggest that our conjecture is correct; see Figure 3 below for an example).

When $\underline{\theta} < \underline{\theta}^{tr}(c)$, the quality of the best predictor increases while the quality of the worst predictor used by speculator decreases when data become more abundant (see Proposition 4). Thus, the range of quality for the predictors used in equilibrium gets wider. This effect increases the dispersion of the quality of predictors used by speculators ($\mathbb{V}[\tau(\theta)]$ increases),

which increases the dispersion of speculators' expected profits, holding price informativeness constant. In equilibrium, price informativeness improves but for $\underline{\theta}$ small enough, this second effect is not sufficient to offset the first. This explains the second part of the proposition.

The effect of a decrease in α is more straightfoward. Indeed, such a decrease leads speculators to be less demanding for the quality of their predictors (θ^* increases when α decreases). Thus, a decrease in α enlarges the dispersion of the quality of speculators' predictors. As it also reduces price informativeness, it follows from eq.(26) that the dispersion of speculators' trading profits increases.

In sum, data abundance and improvements in computing power have similar effects on speculators' expected profits but can have opposite effects on the dispersion of these profits. Figure 3 illustrates this point using the same specifications for the density of θ as in Figure 1. For these specifications, a decrease in the cost of processing data always reduces the dispersion of expected trading profits across speculators. In contrast, the dispersion expected trading profits always increases when $\underline{\theta}$ decreases (for the numerical values considered in Figure 3).

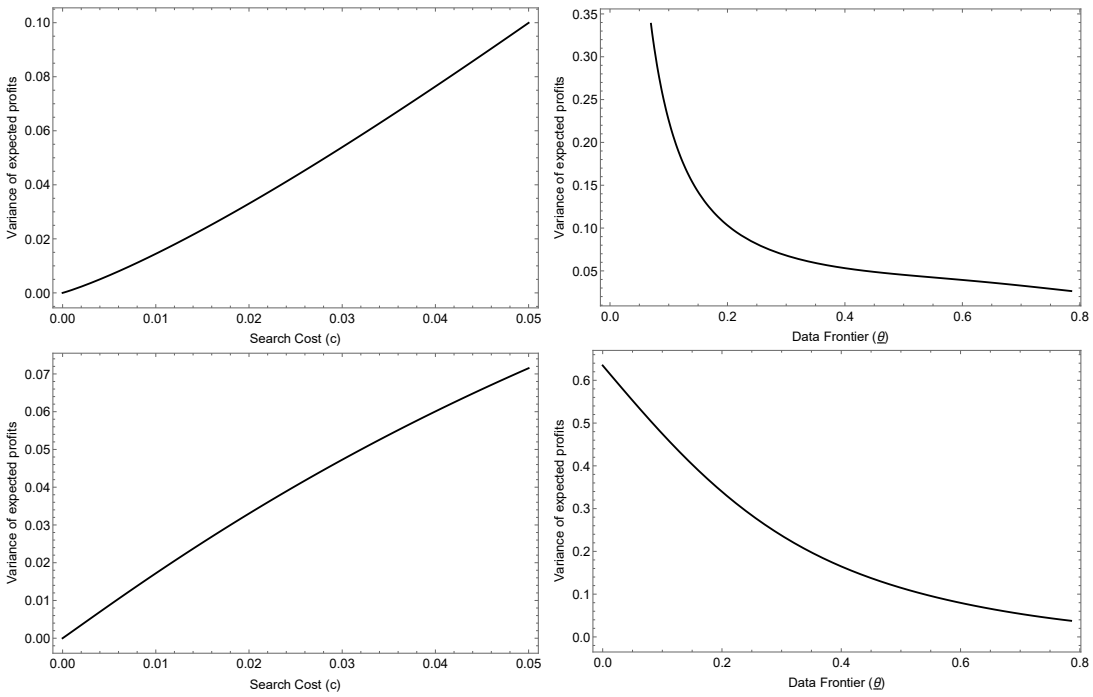

Figure 3: The left hand-side graph represents the variance of speculators expected profits, $\mathbb{V}[\pi(\theta)]$, as a function of the search cost, c, with other parameter values, $\underline{\theta} = \pi/5, \rho = 1, \sigma^2 = 1, \nu^2 = 1$. The right hand-side graph represents the variance of speculators expected profits as a function of the data frontier, $\underline{\theta}$, with other parameter values, $c = 0.05, \rho = 1, \sigma^2 = 1, \nu^2 = 1$. In the first figure, we assume that $\phi(\theta) = 3cos(\theta)sin^2(\theta)$ (Case 1) while in the second we assume that $\phi(\theta) = 5cos(\theta)sin^4(\theta)$ (Case 2).

4.3.1 Data Abundance, Computing Power and Crowding

Crowding, i.e., the tendency for investors to follow the same trading strategy and exploit the same signals, has been a rising concern for quantitative fund managers. Shanta Putchler, the CEO of Mannumeric (a quantitative investment fund) notes that:[14] "*The single largest contributor to crowding is the simple fact that investors tend to do the same sorts of things. There is a real propensity for investors to analyse the same datasets, with the same statistical techniques, and hence end up with largely overlapping positions.*" An interesting question is whether data abundance and greater computing power reduces or increases crowding.

To analyze this question, we study how data abundance and computing power affect the correlation between speculators' equilibrium positions (a measure of crowdness of trading strategies in a given asset) in our model. Specifically, let $cov(x(s_{\theta_i}, p^*), x(s_{\theta_j}, p^*))$ be the covariance between the equilibrium holdings of a speculator with type θ_i and a speculator with type θ_j. Using eq.(23), we obtain:

$$cov(x^*(s_{\theta_i}, p^*), x^*(s_{\theta_j}, p^*)) = \frac{\tau(\theta_i)\tau(\theta_j)}{\sigma^4\rho^2}\mathbb{V}[\omega - p] = \frac{\tau(\theta_i)\tau(\theta_j)}{\sigma^4\rho^2\mathcal{I}(\theta^*, \underline{\theta})}. \tag{27}$$

We deduce that the pairwise correlation between the equilibrium positions of a speculator with type θ_i and a speculator with type θ_j is:

$$corr(x^*(s_{\theta_i}, p^*), x^*(s_{\theta_j}, p^*)) = \left(1 + \frac{\mathcal{I}(\theta^*, \underline{\theta})}{\tau(\theta_i)\tau_\omega}\right)^{-\frac{1}{2}} \left(1 + \frac{\mathcal{I}(\theta^*, \underline{\theta})}{\tau(\theta_j)\tau_\omega}\right)^{-\frac{1}{2}} \tag{28}$$

Thus, holding the quality of the predictors used by two speculators constant, their positions become less correlated when price informativeness is higher. Intuitively, the reason is that speculators trade on the component of their forecast of the asset payoff that is orthogonal to the price. This component reflects both the component of the fundamental, ω, that is not reflected into the equilibrium price and the noise in speculators' signal. The higher the first component relative to the second, the higher the pairwise correlation in speculators' positions in the asset. As the price becomes more informative, the first component becomes smaller and smaller relative to the noise component and as a result, the pairwise correlation between speculators' positions drops. Using Proposition 7, we deduce the following result.

Proposition 10.

[14] See https://www.man.com/maninstitute/crowding)

1. *Greater computing power (a decrease in c) reduces the pairwise correlation of speculators' positions.*

2. *Data abundance has an ambiguous effect on the pairwise correlation of speculators' positions. It reduces it if it improves price informativeness but increases it otherwise.*

This proposition suggests again that data abundance and computing power do not necessarily have the same effects. To test the previous result, one must be able to measure the pairwise correlation of speculators' positions, holding the quality of their signal constant. One difficulty is that data abundance or greater computing power can change the whole distribution of speculators' predictors used in equilibrium, in particular the mean and variance of this distribution (see Propositions 8 and 9). Intuitively, a decrease in the dispersion of predictors might on *average* raise the pairwise correlation in speculators' positions. To analyze this possibility, we consider the average pairwise correlation in speculators' positions across all possible pairs, defined as:

$$\overline{corr} \equiv \mathbb{E}\left[corr(x^*(s_{\theta_i}, p^*), x^*(s_{\theta_j}, p^*)) \middle| \theta_i, \theta_j \in [\underline{\theta}, \theta^*]^2 \right] = \mathbb{E}\left[\left(1 + \frac{\mathcal{I}(\theta^*, \underline{\theta})}{\tau(\theta)\tau_\omega} \right)^{-\frac{1}{2}} \middle| \underline{\theta} < \theta < \theta^* \right]^2. \tag{29}$$

The effect of greater computing power or data abundance on this average correlation is a priori ambiguous. To see this, consider a decrease of θ^* induced by the reduction of c. On the one hand, every pairwise correlation decreases as shown in Proposition 10, which drives down the average correlation. But a composition effect goes in the opposite direction. Indeed, the correlations between the former marginal type θ^* and other θ's cease to exists, and they are replaced by the average correlation which is larger. Consistently, Figure 4 shows that the average correlation in speculators' positions can be non monotonic in c or $\underline{\theta}$ (for the same reason).

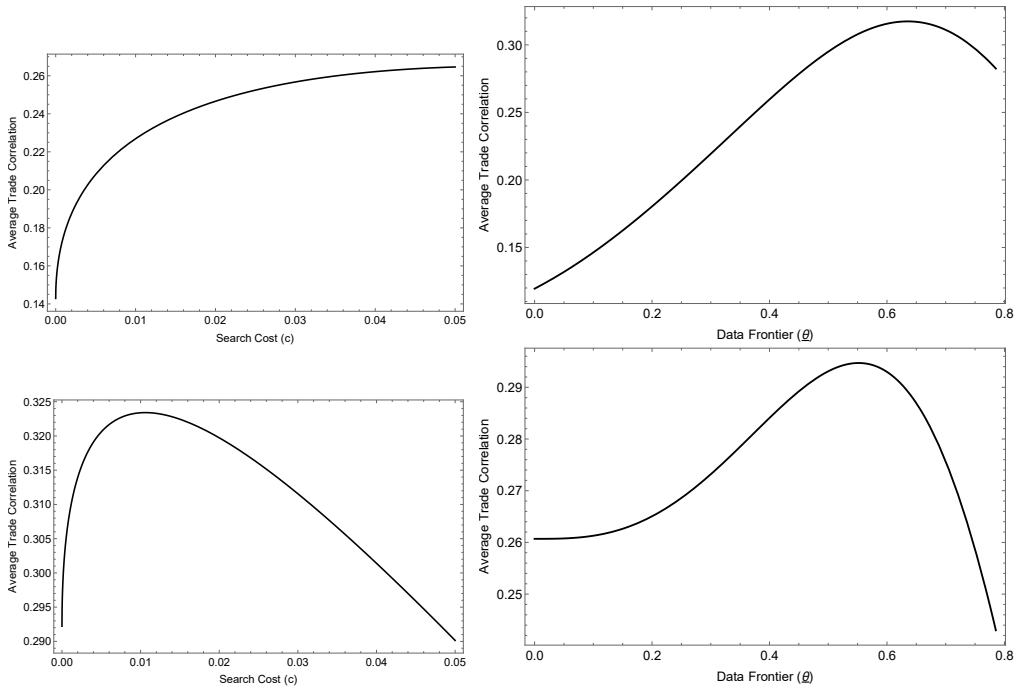

Figure 4: The left hand-side graph represents the average trade correlation, $\overline{corr}$, as a function of the search cost, c, with other parameter values, $\underline{\theta} = \pi/5, \rho = 1, \sigma^2 = 1, \nu^2 = 1$. The right hand-side graph represents the average trade correlation as a function of the data frontier, $\underline{\theta}$, with other parameter values, $c = 0.05, \rho = 1, \sigma^2 = 1, \nu^2 = 1$. In the first figure, we assume that $\phi(\theta) = 3cos(\theta)sin^2(\theta)$ (Case 1) while in the second we assume that $\phi(\theta) = 5cos(\theta)sin^4(\theta)$ (Case 2).

5. Speculators' Welfare and Data Abundance

As shown in Section 3.1, the ex-ante expected utility of each speculator in equilibrium is $J(\theta^*, \theta^*) = g(\theta^*, \theta^*)$. That is, each speculator's expected utility is just equal to the expected utility from trading on the worst predictor used in equilibrium. The reason is that the increase in the expected utility from trading associated with further explorations for a speculator who has found a predictor with type θ^* is just offset by the expected utility cost of further explorations.

As can be seen from eq.(12), $g(\theta^*, \theta^*)$ decreases with the informativeness of the asset price and increases with the quality of the worst predictor ($\tau(\theta^*)$). Now, when $\underline{\theta} < \underline{\theta}^{tr}(c)$, a decrease in $\underline{\theta}$ raises price informativeness (Proposition **??**) and reduces the quality of the worst predictor (Proposition 4). Thus, it unambiguously reduces speculators' expected utility.

Proposition 11. *When $\underline{\theta} < \underline{\theta}^{tr}(c)$, pushing back the data frontier (a decrease in $\underline{\theta}$) reduces speculators' expected utility.*

In contrast, an increase in computing power increases the quality of the worst predictor and

price informativeness. Thus, its effect on speculators' expected utility is ambiguous. Numerical simulations show that the first effect dominates unless c becomes very small. Thus, in contrast to a push back of the data frontier, an improvement in computing power raises speculators' expected utility. Figure 5 illustrates this point using the same numerical examples as in Figure 1.

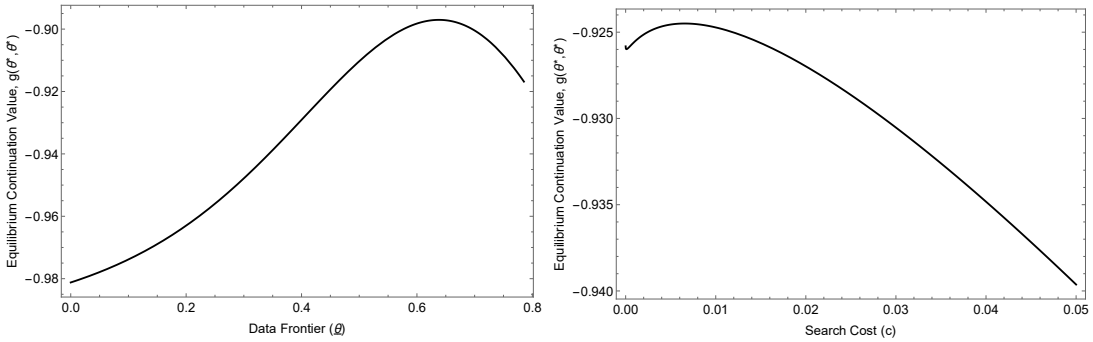

Figure 5: This graph shows speculators' ex-ante expected utility as a function of $\underline{\theta}$ and c when $\phi(\theta) = 3cos(\theta)sin^2(\theta)$ (with other parameter values being set at $\rho = 1, \sigma^2 = 1, \nu^2 = 1$).

Interestingly, data abundance can make speculators worse off in equilibrium. One might then wonder whether it would not be optimal for a speculator to ignore new data. This is not the case, however. To see this, suppose that $\underline{\theta}$ drops from $\underline{\theta_0}$ to $\underline{\theta_1} < \underline{\theta_0}$ but that speculators decide not to take advantage of the new data. This means that they keep behaving as if $\underline{\theta} = \underline{\theta_0}$ (in particular, they use the stopping rule $\theta^*(\underline{\theta_0})$). A speculator's expected utility is then given by $J(\theta^*(\underline{\theta_0}), \theta^*(\underline{\theta_0}))$. One can show (using eq.(14) that, holding $\theta^*(\underline{\theta_0})$ constant, this expected utility increases when the quality of the best predictor is improved. Thus, if new datasets open the possibility that the quality of the best predictor improves, each speculator will individually find optimal to use these datasets, if she expects others not to do so. But then the equilibrium stopping rule must shift from $\theta^*(\underline{\theta_0})$ to $\theta^*(\underline{\theta_1})$

We now show that the equilibrium stopping rule is excessive, in the sense that it leads them to search predictors of too high quality compared to what would maximize speculators' ex-ante expected utility, if speculators could commit to follow to a particular stopping rule. To see this, let θ^{**} be the stopping rule that maximizes the ex-ante expected utility of all speculators, i.e., that solves:

$$\theta^{**} = \arg\max_{\theta} \ \Pi(\theta, \underline{\theta}). \tag{30}$$

We obtain the following result.

Proposition 12. *In equilibrium, the quality of the marginal predictor is too high relative to the quality that maximizes speculators' ex-ante expected utility, that is*

$$\theta^* < \theta^{**}. \tag{31}$$

Thus, there is excessive investment in search in equilibrium from speculators' viewpoint. The reason is as follows. Suppose that all speculators search for predictors with a stopping rule equal to θ^{**}. Now consider a speculator who draws a predictor with precision θ^{**}. Her expected trading profit is less than her expected utility of continuing searching for another predictor, assuming that other speculators keep searching with the same intensity (i.e., the same stopping rule θ^{**}). Formally:

$$g(\theta^{**}, \theta^{**}) < J(\theta^{**}, \theta^{**}).$$

Thus, the speculator has an incentive to deviate from the stopping rule by increasing her search intensity. However, in doing so, the speculator ignores the fact that this must be true for all speculators and that if all speculators deviate, there will all be worse off. Instead, a central planner organizing the search for predictors would internalize this effect.

6. Conclusion

To be written

References

Abis, Simona, 2018, Man vs machine: Quantitative and discretionary equity management, *Working paper* .

Bai, Jennie, Thomas Phillipon, and Alexi Savov, 2016, Have financial markets becomemore informative?, *Journal of Financial Economics* 122, 625–654.

Brogaard, Jonathan, and Abalfazl Zareei, 2019, Machine learning and the stock market, Technical report.

Dugast, Jerome, and Thierry Foucault, 2018, Data abundance and asset price informativeness, *Journal of Financial economics* 130, 367–391.

Farboodi, Maryam, Adrien Matray, and Laura Veldkamp, 2019, Where has all the data gone?, *Working paper* .

Farboodi, Maryam, and Laura Veldkamp, 2019, Long run growth of financial technology, *forthcoming American Economic Review* .

Gao, Meng, and Jiekun Huang, 2019, Informing the Market: The Effect of Modern Information Technologies on Information Production, *The Review of Financial Studies* .

Grennan, Jillian, and Roni Michaely, 2019, Fintechs and the market for financial aanalysis, *Working paper* .

Grossman, Sanford, and Joseph Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393–408.

Han, Jungsuk, and Francesco Sangiorgi, 2018, Searching for information, *Journal of Economic Theory* 175, 342–373.

Harvey, Campbell, 2017, The scientific outlook in financial economics, *Journal of Finance* 72, 1399–1440.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2016, A rational theory of mutual funds' attention allocation, *Econometrica* 84, 571–626.

Katona, Zsolt, Markus Painter, Panos Patatoukas, and JienYin Zengi, 2019, On the capital market consequences of alternative data: Evidence from outer space, Technical report.

Marenzi, Octavio, 2017, Alternative data: The new frontier in asset management, *Report, Optimas Research* .

Nordhaus, William, 2015, Are we approaching an economic singularity? information technology and the future of economic growth, *Yale University Cowles Foundation Discussion Papers 2021* .

Shyang Huang, Yang Xiong, and Liyan Yang, 2020, Information skills and data sales, *Working paper* .

Veldkamp, Laura, 2011, Information choice in macroeconomics and finance .

Veldkamp, Laura, and Cindy Chung, 2020, Data and the aggregate economy, *Forthcoming Journal of Economic Literature* .

Verrecchia, Robert, 1982, Information acquisition in a noisy rational expectations economy, *Econometrica* 1415–1430.

Vives, Xavier, 1995, Short-term investment and the informational efficiency of the market, *Review of Financial Studies* 8, 125–160.

Yan, Xuemin (Sterling), and Lingling Zheng, 2017, Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach, *The Review of Financial Studies* 30, 1382–1423.

Zhu, Christina, 2019, Big data as a governance mechanism, *Review of Financial Studies* 32, 2021–2061.

A. Proofs

To be written