Data Abundance and Asset Price Informativeness*

Jérôme Dugast† Thierry Foucault‡

March 25, 2017

Abstract

Information processing filters out the noise in data but it takes time. Hence, low precision signals are available before high precision signals. To capture this feature, we develop a model of securities trading in which investors can acquire signals (about future cash flows) of increasing precision over time. As the cost of producing low precision signals declines, prices are more likely to reflect these signals before more precise signals become available. This effect increases price informativeness in the short run but not necessarily in the long run, because it reduces the profit from trading on more precise signals. We make additional predictions for trade and price patterns.

KEYWORDS: Asset Price Informativeness, Big Data, FinTech, Information Processing, Markets for Information, Contrarian and momentum trading.

*This paper is a significantly revised version of our paper previously entitled "False news, informational efficiency, and price reversals." We thank Torben Andersen, Olivier Dessaint, Eugene Kandel, Terrence Hendershott, Johan Hombert, Mark Lipson, Stefano Lovo, Maureen O'Hara, Katya Malinova, Albert Menkveld, Sophie Moinas, Christine Parlour, Emiliano Pagnotta, Rafael Repullo, Patrik Sandas, Javier Suarez, Pietro Veronesi, Laura Veldkamp, Albert Wang, Eric Weisbrod, Brian Weller, and Marius Zoican for useful comments and suggestions. We are also grateful to conference and seminar participants at the 2017 AFA Meetings in Chicago, the 2016 High Frequency trading Workshop in Vienna, the 9^{th} hedge fund conference in Paris Dauphine, the CEPR first annual spring symposium in Financial Economics at Imperial college, the CEMFI, the conference on Banking and Finance in Portsmouth, the Jan Mossin Memorial Symposium, the 2014 SFS Finance Cavalcade, the 2014 European Finance Association Meetings, the Banque de France International Workshop on Algorithmic Trading and High Frequency Trading, the Board of Governors of the Federal Reserve System, the U.S. Securities and Exchange Commission, the University of Virginia, and the University of Maryland for their feedback. This work is funded by a grant from the Institut Louis Bachelier. Thierry Foucault also acknowledges financial support from the Investissements d'Avenir Labex (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047).

†Luxemburg school of finance. Tel: (+352) 46 66 44 5212; E-mail: jerome.dugast@uni.lu
‡HEC, Paris and CEPR. Tel: (33) 1 39 67 95 69; E-mail: foucault@hec.fr

"Increasingly, there is a new technological race in which hedge funds and other well-heeled investors armed with big data analytics analyze millions of twitter messages and other non-traditional information sources to buy and sell stocks faster than smaller investors can hit "retweet"."

in "How investors are using social media to make money," Fortune, December 7, 2015.

1 Introduction

Improvements in information technologies change how information is produced and disseminated in financial markets. In particular, they enable investors to obtain huge amount of data at lower costs.[1] For instance, investors can now easily get on-line access to companies reports, economic reports, or investors' opinions (expressed on social medias) to assess the value of a stock.[2] Similarly, traditional data vendors like Reuters, Bloomberg, or "Fintechs" (e.g., iSentium, Dataminr, or Eagle Alpha) use so-called news analytics to extract signals from unstructured data (news reports, press releases, stock market announcements, tweets, satellite images etc.) and sell these signals to investors who feed them into their trading algorithms.[3]

How does this evolution affect the informativeness of asset prices? This question is important because, ultimately, more informative prices enhance the efficiency of capital allocation (see Bond, Edmans, and Goldstein (2012) for a survey). Economists should a-priori expect the decline in the cost of accessing information to enhance asset price informativeness. Indeed, extant models with endogenous information acquisition predict that asset price informativeness increases when information costs decline, either because more investors buy information (Grossman and Stiglitz (1980)) or because investors acquire more precise signals (Verrechia (1982)).

However, being static, these models ignore the time dimension of information production, namely, filtering out noise from data takes time. For this reason, less precise signals become available before more precise signals, not the other way round. This timing is

[1]At the turn of the millennium, the cost of sending one trillion bits was already only $0.17 (versus $150,000 in 1970); see "The new paradigm", Federal Reserve Bank of Dallas, 1999.

[2]For instance, websites such as StockTwits or Seeking Alphas allow investors to comment on stocks, share investment ideas, and provide, in real time, raw financial information pulled off from other social medias. For evidence that information exchanged on social medias contains value relevant information, see Chen et al.(2014).

[3]For popular press articles on this evolution, see, for instance, "*Rise of the news-reading machines*" (Financial Times, January 26, 2010); "*How investors are using social medias to make money*" (Fortune, December 7, 2015); "*Investors mine big data for cutting hedge strategies*" (Financial Times, March 30, 2016); or "*Big data is a big mess for hedge funds hunting signals*" (Bloomberg, November 22, 2016).

particularly relevant when considering on-line data such as tweets, newswires, companies reports etc., which often are very noisy and requires accumulation of more data to generate precise signals. Thus, early signals produced with such data (e.g., using machine learning or natural language processing) are less precise than later signals obtained with a deeper analysis.[4] In this paper, we show that, for this reason, a decline in the cost of accessing data can *reduce* the long run informativeness of asset prices about fundamentals, in contrast to the prediction of existing models of information acquisition.

In our model, information sellers produce a "raw" (i.e., unfiltered) signal and a "processed" (i.e., filtered) signal about the payoff of a risky asset and sell these to speculators (for a fee specific to each signal). The raw signal is correct (reveals the asset payoff) with probability θ or is just noise with probability $(1-\theta)$. Thus, θ characterizes the reliability of the raw signal. The true type of the signal (information/noise) can only be discovered after filtering out the noise from raw data, which requires some time. To account for this delay, we assume that the processed signal (i.e., the raw signal *and* its type) is available with a lag relative to the raw signal. Specifically, the raw signal is available in period 1 while the processed signal is only available in period 2.[5] Thus, speculators who buy the processed signal trade with a lag relative to speculators who buy the raw signal. When they receive their signal, speculators can trade on it with a risk neutral market maker and liquidity traders (as in Kyle (1985)).

Following Veldkamp (2006a,b), we assume that the costs of producing the raw and the processed signals are fixed but, once produced, each signal can be replicated for free (the marginal cost of providing a signal to an extra user is zero). Furthermore, the market for information is competitive: (i) raw and processed signals are sold at competitive fees (i.e.,

[4] In line with this idea, examples of erroneous trading decisions and large price changes due to misleading signals based on web data abound. For instance, on April 23, 2013 a fake tweet from a hacked Associated Press twitter account announced that explosions at the White House had injured President Barack Obama. The Dow Jones immediately lost 145 basis points but it recovered in less than three minutes after the news proved to be false. Commenting this event, the Economist Magazine writes: *"human users must extract some sort of signal every day from the noise of innumerable tweets. Computerised trading algorithms that scan news stories for words like "explosions" may have proved less discerning and triggered the sell-off. That suggests a need for more sophisticated algorithms that look for multiple sources to confirm stories."* (see The Economist, "#newscrashrecover", April 27, 2013.). See also "How investors are using social media to make money," Fortune, December 7, 2015 for other examples in the same vein.

[5] One possible interpretation of this timing is as follows. News (e.g., a new SEC filing or an earning conference call by a firm) about the asset cash flows arrives just before date 1. Based on this news, information sellers produce signals about the asset. For instance, the raw signal might be a buy/sell recommendation for the asset based on linguistic analysis of the news regarding the asset while the processed signal is a buy/sell recommendation based on a deeper analysis (e.g., financial statements analysis) of the implications of the news for the asset value. As the processed signal requires human intervention and accumulation of more data, it takes more time to produce than the raw signal.

such that information sellers make zero profit) and (ii) speculators' expected profit from trading on one signal net of information fees is equal to zero. In this set-up, we analyze how a decline in the cost of producing the raw signal affects equilibrium outcomes, in particular the equilibrium demands for each signal (i.e., the number of speculators buying it) and the informational content of the asset price in the short run (period 1) and the long run (period 2).

We first show that a decrease in the cost of producing the *raw* signal can *strengthen* or *reduce* the demand for the *processed* signal in equilibrium. Indeed, this decrease raises the number of speculators who trade on the raw signal and therefore the likelihood that the price of the asset reflects this signal in period 1, i.e., before speculators receive the processed signal. When the raw signal is just noise, this effect is beneficial for those who buy the processed signal. Indeed, they learn from their signal that the asset is mispriced, due to the noise injected in the price by those who trade on the raw signal, and they can exploit this mispricing. In contrast, when the raw signal is not noise, a more informative price at date 1 makes those who buy the processed signal worse off since it reduces their informational advantage relative to the market maker.

Thus, the net effect of a decrease in the cost of producing the raw signal on the value of the processed signal and therefore the demand for this signal is ambiguous. It strengthens the demand for the processed signal in equilibrium (i.e., after accounting for the adjustment in the fees charged by information sellers to the decrease in the cost of the raw signal) only if the raw signal is sufficiently unreliable (i.e., $\theta < \hat{\theta} < 1/2$). Otherwise, a decrease in the cost of producing the raw signal *reduces* the demand for the processed signal in equilibrium, as if "bad" signals were driving out "good" signals. In this case, a decline in the cost of the raw signal makes prices more informative in the short run and yet, paradoxically, less informative in the long run.[6]

This implication of the model is consistent with Weller (2016) who finds empirically a negative association between the activity of algorithmic traders (a class of traders who is likely to trade on relatively raw signals) and the informativeness of prices about future earnings. It also offers a possible interpretation of the empirical findings in Bai, Phillipon, and Savov (2015). For the entire universe of U.S. stocks, they find (see their Figure A.3) that stock price informativeness has been declining over time (they obtain the opposite conclusion for stocks in the S&P500 index). They attribute this trend to change in the

[6]There is at, least as much information available in period 2 than in period 1 and strictly more if, in equilibrium, some speculators buy the processed signal. Thus, the informativeness of the price in period 2 is weakly higher than in period 1. Yet, when $\theta > \hat{\theta}$ and the cost of producing the raw signal decreases, the informativeness of the price in period 2 decreases, even though it increases in period 1.

characteristics of public firms in the U.S. Our model suggests that the reduction in the cost of producing raw signals might be another explanatory factor for this evolution.

Our model has additional testable implications for the trade patterns of various types of investors. First the model predicts that the correlation between the order flow (the difference between buys and sells) of speculators trading on the raw signal and that of speculators trading on the processed signal should decline (and could even become negative) when the cost of producing the raw signal decreases. Indeed, speculators receiving the processed signal trade in a direction opposite to that of speculators who trade on the raw signal when (i) the raw signal is noise and (ii) it is reflected in the price before speculators receive the processed signal. Holding the precision of the raw signal (θ) constant, this event occurs more frequently when more speculators trade on the raw signal, i.e., when its production cost is small. For this reason, when this cost declines, sell orders (resp., buy orders) from speculators who trade on the raw signal are more likely to be followed by buy orders (resp., sell orders) from speculators who trade on the processed signal.

Second, the order flow from speculators who trade on the processed signal and past returns are correlated.[7] This correlation is negative when the raw signal is sufficiently unreliable ($\theta \leq \frac{1}{2}$) and positive otherwise. Intuitively, price changes due to trades exploiting the raw signal are more likely to be due to noise, and therefore subsequently corrected by speculators who receive the processed signal, when the reliability of the raw signal, θ, is low enough. Thus, in equilibrium, speculators who trade on the processed signal behave either like *contrarian* traders (they trade against past returns) when the raw signal is unreliable or *momentum* traders (they trade in the same direction as past returns) when the raw signal is more reliable. The model also implies that, in absolute value, the correlation between the order flow from speculators who trade on the processed signal and past returns should be weaker when the cost of producing the raw signal declines.

Last, the direction of the order flow from speculators who trade on the raw signal is positively correlated with future returns.[8] However, this correlation becomes weaker when the cost of producing the raw signal declines. Indeed, this decline increases the demand for the raw signal and therefore the likelihood that the asset price fully reflects

[7]This prediction is non standard. Indeed, standard models of informed trading (e.g., Kyle (1985)) predicts a zero correlation between the trades of informed investors at a given date and lagged returns (see Boulatov, Livdan and Hendershott (2012), Proposition 1, for instance.)

[8]This is not due to serial correlation in returns. In our model, the price of the asset at each date is equal to its expected value conditional on all available public information, i.e., the history of trades as in Kyle (1985). Hence, returns are serially uncorrelated in our model.

this signal before the arrival of the processed signal. When this happens, speculators who receive the processed signal can only profitably trade on the component of their signal that is orthogonal to the raw signal. This effect reduces the predictive power of the order flow from speculators trading on the raw signal about the order flow from speculators trading on the processed signal, and therefore future returns.

All our predictions are about the effects of a decline in the cost of producing raw signals. Empiricists could test these predictions by using shocks to the cost of accessing raw financial data. For instance, in 2009, the SEC mandated that financial statements be filed with a new language (the so called EXtensible Business Reporting Language or XBLR) on the ground that it would lower the cost of accessing data for smaller investors.[9] The implementation of this new rule or other similar shocks could therefore be used to test some of our predictions.[10]

We discuss the literature related to our paper in the next section. Section 3 describes the model. Section 4 derives equilibrium prices at dates 1 and 2, taking the demands for the raw and the processed signals as given while Section 5 endogenizes these demands. Section 6 derives the implications of the model for (i) asset price informativeness and (ii) price and trade patterns. Section 7 presents an extension of the model in which investors can make their decision to buy the processed signal contingent on the short run evolution of asset prices. Section 8 concludes. Proofs of the main results are in the appendix. Additional material is provided in a companion appendix available on the authors' website.

2 Related Theoretical Literature

Our paper contributes to the literature on costly information acquisition in financial markets (e.g., Grossman and Stiglitz (1980), Verrechia (1982), Admati and Pfleiderer (1986), Veldkamp (2006a,b), Cespa (2008), or Lee (2013); see Veldkamp (2011) for a

[9] See SEC (2009). In particular on page 129, the SEC notes that: "*If [XBLR] serves to lower the data aggregation costs as expected, then it is further expected that smaller investors will have greater access to financial data than before. In particular, many investors that had neither the time nor financial resources to procure broadly aggregated financial data prior to interactive data will have lower cost access than before interactive data. Lower data aggregation costs will allow investors to either aggregate the data on their own, or purchase it at a lower cost than what would be required prior to interactive data. Hence, smaller investors will have fewer informational barriers that separate them from larger investors with greater financial resources.*"'

[10] Interestingly, data vendors such as Dow Jones screen SEC filings by firms and release information contained in these filings through specialized services (e.g., Dow Jones Corporate Filing Alert). Thus, a reduction in the cost of accessing these filings for data vendors is similar to a reduction in the cost of producing the raw signal in our model.

survey). Some papers in this literature (e.g., Verrechia (1982) or Peress (2010)) have considered the case in which investors can choose the precision of their signals, assuming that the cost of a signal increases with its precision.

Our model differs from these models in two important dimensions. First, in the extant literature, investors trade on their signals simultaneously while in our model, traders who buy a signal with relatively low precision can trade *before* those who buy a signal with higher precision, because information processing takes time. Second, in extant models, the cost of precision is exogenously specified while it is endogenous in our model. In particular, in equilibrium, the fee charged for the high precision signal (its acquisition cost for investors) indirectly depends on the demand for the low precision signal and thereby the production cost of this signal. As a result, we obtain different implications, which highlights the importance of the time dimension in producing more precise signals. In particular, in Verrechia (1982), a decline in the cost of precision increases price informativeness (see Verrechia (1982), Corollary 4), whether this decrease regards high or low precision signals (or both). In contrast, in our model, a decrease in the cost of producing the low precision (raw) signal can *reduce* price informativeness.

In Lee (2013), investors can buy signals about one of two independent fundamentals for an asset, say, A, and B. As all informed investors trade simultaneously, an increase in the number of speculators trading on, say, A blurs market makers' ability to learn (from the order flow) about B while raising the price impact of all market orders. Interestingly, the first effect enhances the expected profit of speculators who trade on B while the second decreases it. For this reason, in Lee (2013), the number of investors informed about one fundamental can increase or decrease with the number of investors informed about the other one.

Similarly, in our model, the mass of investors informed about the processed signal can increase or decrease with the mass of speculators informed about the raw signal. However, the economic mechanisms in our model are different from those in Lee (2013). Indeed, in our set-up, traders with different signals trade at different dates. Thus, their orders are not batched and therefore do not have the same price impact. Moreover, speculators' signals are correlated in our model. For this reason, an increase in the mass of speculators trading on the raw signal at date 1 allows market makers to draw *more*, not less, precise inferences from the order flow at this date about both the asset payoff *and* the signal subsequently received by those trading on the processed signal.

Holden and Subrahmanyam (1996) and Brunnermeier (2005) consider models with two trading rounds (dates 1 and 2) in which investors receive signals about two independent risk factors affecting the payoff of a risky asset. One factor is publicly released at date 2 while the other remains unknown until the asset pays off (date 3). In Holden and Subrahmanyam (1996), investors are risk averse and can choose to trade either on a signal about the factor revealed at date 2 (the short term signal) or a signal about the factor revealed at date 3 (the long term signal). They show that more investors choose to be informed about the short term signal when risk aversion increases. In our model, investors are risk neutral and can buy both signals. Changes in the demand for a signal are driven by variation in the cost of producing the raw signal and our novel predictions pertain to changes in this cost. Brunnemeier (2005) shows that the asset price is more informative at date 2 than when no investor is informed about the factor revealed at date 2. In contrast to Brunnemeier (2005), our results about price informativeness are not driven by speculation on forthcoming public news. Instead, they reflect a change in the relative demands for signals of low and high precisions (the number of informed investors is exogenous in Brunnermeier (2005)).

As in Froot, Scharfstein and Stein (1992) and Hirshleifer et al.(1994), our model features "early" (those who trade on the raw signal) and "late" (those who trade on the processed signal) informed investors. In these models, the number of early and late informed traders is exogenous. In contrast, in our model, the number of traders trading on the early (raw) signal or the late (processed) signal is endogenous and late traders have signals of higher precision.[11] For these reasons, the implications of our model are distinct from those in other models with early and late informed investors.[12] For instance, in Hirshleifer et al.(1994), the trades of early and late informed investors are always positively correlated (see their Proposition 2) while, instead, they can be negatively correlated in our model. Moreover, in Hirshleifer et al.(1994), the trades of late informed investors are not correlated with past returns (see their Proposition 3) while they are in our model.

[11] In Holden and Subrahmanyam (2002), risk averse investors can choose to receive information at dates 1 or 2. However, the precision of investors' signals is the same at both dates. In contrast, in our model, late informed investors receive a signal of higher precision.

[12] In Froot et al.(1992), there exist equilibria in which a fraction of speculators trade on noise. However, there is no possibility for traders to correct price changes due to such trades. In contrast, in our model, speculators correct price changes due to noise, after receiving the processed signal.

3 Model

We consider the market for a risky asset. Figure 1 describes the timing of actions and events. There are four periods ($t \in \{0, 1, 2, 3\}$). The payoff of the asset, V, is realized at date $t = 3$ and can be equal to 0 or 1 with equal probabilities. Trades take place at dates $t = 1$ and $t = 2$ among: (i) a continuum of liquidity traders, (ii) a continuum of risk neutral speculators, and (iii) a competitive and risk neutral market-maker. We denote by $\bar{\alpha}$ the mass of speculators relative to the mass of liquidity traders (which we normalize to one). At date 0, speculators can buy signals about the payoff of the asset. As explained below, these signals are delivered by information sellers at date 1 or 2, depending on their type.

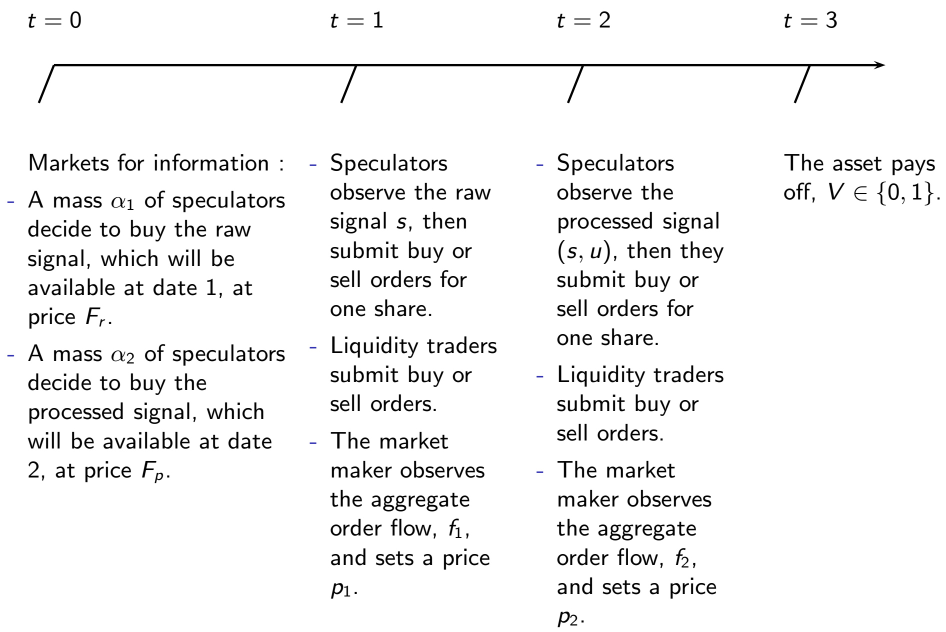


Figure 1: Timing

The raw and the processed signals. Just before date 1, new data about the payoff of the asset becomes available. These data are the raw material used by information sellers to produce two types of signals: (i) an unfiltered signal (henceforth the "raw signal") and (ii) a filtered signal (henceforth the "processed signal"). The raw signal, $\tilde{s}$, is:

$$\tilde{s} = \tilde{u} \times \tilde{V} + (1 - \tilde{u}) \times \tilde{\epsilon}, \tag{1}$$

where $\tilde{u}$, $\tilde{\epsilon}$, and $\tilde{V}$ are independent and can be equal to 0 or 1. Specifically, $\tilde{u} = 1$ with probability θ while $\tilde{\epsilon} = 1$ with probability 1/2. Thus, with probability θ, the raw signal is equal to the asset fundamental while with probability $(1-\theta)$, it is just noise.[13]

The processed signal is obtained after filtering out the noise from the new data available just before date 1 (e.g., by accumulating more data). Thus, the processed signal is the pair (s, u), that is, the raw signal and its type (noise or fundamental). We say that the processed signal "confirms" the raw signal if $u = 1$ and "invalidates" it if $u = 0$. For the problem to be interesting, we assume that $0 < \theta < 1$ so that the raw signal is informative but less reliable than the processed signal.

To capture the idea that information processing takes time, we assume that producing the processed signal requires one more period than producing the raw signal. Thus, the raw signal is delivered (by sellers of this signal) at date $t = 1$ while the processed signal is delivered at date $t = 2$.

The market for information is opened at date 0. That is, at this date, information sellers set their fee for each type of signal and speculators decide to subscribe or not to their services. Each speculator can choose to (i) buy both types of signals, (ii) only one type, or (iii) no signal at all. The mass of speculators buying the signal available at date t is represented by α_t.[14] We denote by F_r and F_p the fees charged at date 0, respectively, by the sellers of the raw signal and the sellers of the processed signal. We analyze how they are related to the cost of producing each signal in Section 5 when we endogenize these fees.

The asset market. Trading in the market for the risky asset takes place at dates 1 and 2. As in Glosten and Milgrom (1985), each speculator can buy or sell a fixed number of shares–normalized to one–using market orders (i.e., orders that are non contingent on the contemporaneous execution price). If he decides to trade, a speculator will optimally submit an order of the maximum size (one share) because he is risk neutral and too small to individually affect the equilibrium price. To simplify the analysis, we assume that speculators who only buy the raw signal trade at date 1 but not at date 2 (traders who buy both signals can trade at both dates).[15]

[13] The raw signal does not need to be construed as being completely unprocessed data. For instance, firms (e.g., Reuters, Bloomberg, Dataminr, Thinknum, Orbital Insights etc.) selling signals extracted from social medias (twitter etc.), companies reports, or satellite imagery use algorithms to process raw data to some extent. This processing is faster but not as deep as that performed by securities analysts or investment advisors who take the time to accumulate more information (e.g., by meeting firms' managers, forecast future cash flows, compute discount rates etc.) in order to sharpen the accuracy of their signals.

[14] As the mass of speculators is $\bar{\alpha}$, we have $\alpha_t \leq \bar{\alpha}$ for $t \in \{1, 2\}$. In a previous version of the paper, we considered the case in which each speculator could buy only one type of signal. Results in this case are identical to those obtained when we allow each speculator to buy both signals.

[15] This no retrade constraint for speculators who only buy the raw signal can be justified by the fact

We denote by $x_{it} \in \{-1, 0, 1\}$, the market order submitted by speculator i trading at date t, with $x_{it} = 0$ if speculator i chooses not to trade and $x_{it} = -1$ (resp., $+1$) if he sells (resp., buys) the asset. We focus on equilibria in pure strategies in which all speculators play the same strategy at a given date (symmetric equilibria).[16] Hence, we drop index i when referring to the strategy of a speculator since, at a given date t, all speculators follow the same strategy.

At each date t, liquidity traders buy or sell one share of the asset for exogenous reasons. Their aggregate demand at date t, denoted $\tilde{l}_t$, has a uniform distribution (denoted $\phi(\cdot)$) on $[-1, 1]$ and $\tilde{l}_1$ is independent from $\tilde{l}_2$. Liquidity traders ensure that the order flow at date t is not necessarily fully revealing (see below), which is a pre-requisite for speculators to buy signals (e.g., as in Grossman and Stiglitz (1980)).

At date t, the market-maker absorbs the net demand (the "order flow") from liquidity traders and speculators at a price, p_t, equal to the expected payoff of the asset conditional on his information. As the market-maker does not observe $\tilde{s}$ and $\tilde{u}$ until $t = 3$, the price at date t only depends on the order flow history until this date (as in Kyle (1985)). Formally, let f_t be the order flow at date t:

$$f_t = \tilde{l}_t + \int_0^{\alpha_t} x_{it} di. \tag{2}$$

The asset price at date t is:

$$p_t = \mathsf{E}[\tilde{V}|\Omega_t] = \Pr[\tilde{V} = 1|\Omega_t], \tag{3}$$

where Ω_t is the market-maker's information set at date t ($\Omega_1 = \{f_1\}$ and $\Omega_2 = \{f_2, f_1\}$). At date 0, the asset price is $p_0 = \mathsf{E}(V) = 1/2$. The highest and smallest possible realizations of the order flow at date t are $f_t^{max} \stackrel{\text{def}}{=} (1 + \alpha_t)$ (all investors are buyers at date t) and $f_t^{min} \stackrel{\text{def}}{=} -(1 + \alpha_t)$ (all investors are sellers at date t).

We solve for the equilibrium of the model backward. That is, in the next section, we present speculators' optimal trading strategies and equilibrium prices at dates 1 and 2, for

that their positions are riskier (their profit has a larger variance). Thus, in reality, they are likely to have more stringent position limits than traders who also buy the processed signal. In any case, the no retrade constraint is innocuous when the price at date 1 reveals the raw signal since, in this case, retrading on this signal cannot be optimal. If the price at date 1 does not reveal the raw signal, retrading on this signal at date 2 might sometimes be optimal. Allowing for this possibility however makes the analysis of the equilibrium at date 2 more complex without adding insights.

[16] This restriction is innocuous because there are no other equilibria than symmetric equilibria in pure strategies when speculators' expected profits, gross of the fees paid for the signal, are strictly positive. This condition is necessarily satisfied when α_1 is endogenous because no speculator would buy a signal if his gross expected trading profit is zero (see Lemma 2 in Section 5).

given values of α_1 and α_2. This allows us to compute the ex-ante (date 0) expected profits from trading on each type of signal. Armed with this result, we derive the equilibrium of the market for information in Section 5, that is, the equilibrium fees (F_r and F_p) charged by information sellers and the equilibrium demand (i.e., α_1^e and α_2^e) for each type of signal. We then study (in Section 6) how a reduction in the cost of producing the raw signal affects the demands for each signal and asset price informativeness in equilibrium.

4 Equilibrium Trading Strategies and Prices

Let $\mu(s)$ be expected payoff of the asset at date 1 conditional on signal $s \in \{0, 1\}$. We have:

$$\mu(s) = \mathsf{E}[V|\tilde{s} = s] = \Pr[V = 1|\tilde{s} = s].$$

Hence:

$$\mu(1) = \frac{1+\theta}{2} > \frac{1}{2} \quad \text{and} \quad \mu(0) = \frac{1-\theta}{2} < \frac{1}{2}.$$

At date 1, speculators who buy the raw signal observes s. Thus, we denote their trading strategy by $x_1(s)$ and their expected profit per capita conditional on the realization of the raw signal is:

$$\pi_1(\alpha_1, s) = x_1(s)(\mu(s) - \mathsf{E}[p_1|\tilde{s} = s]).$$

The next proposition provides the equilibrium of the market for the risky asset at date 1 and the ex-ante (date 0) expected trading profit for speculators who buy the raw signal.

Proposition 1. *Let* $\omega(x, \alpha_1) = \frac{\phi(x-\alpha_1)}{\phi(x-\alpha_1)+\phi(x+\alpha_1)}$. *The equilibrium at date* 1 *is as follows:*

1. *Speculators receiving the raw signal buy the asset if* $s = 1$ *and sell it if* $s = 0$ ($x_1(0) = -1$ *and* $x_1(1) = 1$). *Other speculators do not trade.*

2. *The asset price is:*

$$p_1^*(f_1) = \mathsf{E}[\tilde{V}|\tilde{f}_1 = f_1] = \omega(f_1, \alpha_1)\mu(1) + (1 - \omega(f_1, \alpha_1))\mu(0), \tag{4}$$

   for $f_1 \in [f_1^{min}, f_1^{max}]$.

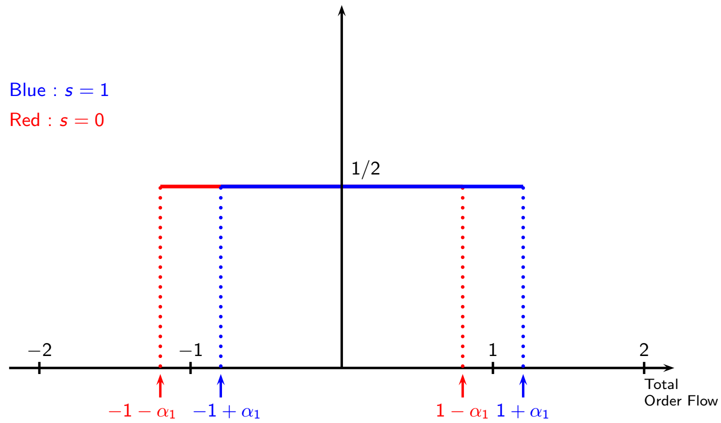
3. *Thus, the ex-ante expected profit from trading on the raw signal is:*

$$\bar{\pi}_1(\alpha_1) \stackrel{def}{=} \mathsf{E}(\pi_1(\alpha_1, s)) = \frac{\theta}{2}\max\{1 - \alpha_1, 0\}. \tag{5}$$

Figure 2: Equilibrium at date 1

Panel A shows the distribution of the order flow at date 1. Panel B shows the equilibrium price at date 1 for each possible realization of the order flow.

(A) Distribution of the Order Flow at Date $t = 1$ (f_1)



(B) Equilibrium Price at Date $t = 1$ (p_1)

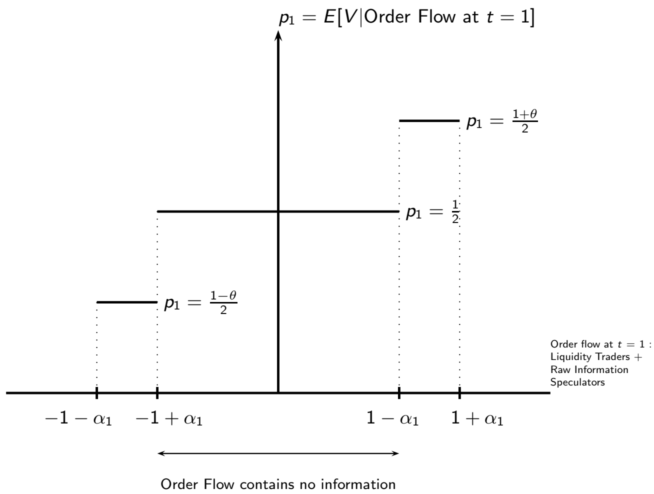


Figure 2 illustrates the proposition. Panel A shows the equilibrium distribution of the aggregate order flow at date 1 for each realization of s, given speculators' trading strategy at this date (remember that the density of liquidity traders' aggregate order, $\phi(.)$, is uniform). Panel B shows the equilibrium price of the asset for each realization of the order flow at date 1. When $s = 0$, speculators who receive the raw signal sell the asset. Thus, their aggregate order is $-\alpha_1$ and the largest possible realization of the order flow in this case is $(1-\alpha_1)$ (when liquidity traders' aggregate order is equal to 1). Thus, when the order flow at date 1 exceeds $(1-\alpha_1)$, the market maker infers that $s=1$ and sets a price equal to $\mu(1)$ (see Panel B in Figure 2). Symmetrically, if the order flow at date 1 is smaller than $-(1-\alpha_1)$, the market maker infers that $s=0$ and sets a price equal to $p_1=\mu(0)$. Intermediate realizations of the order flow at date 1 (those in $[-1+\alpha_1, 1-\alpha_1]$) have the same likelihood whether $s=1$ or $s=0$ (see Panel A in Figure 2). Thus, they provide no information to the market marker and, for these realizations, the market maker sets a price equal to the ex-ante expected value of the asset, 1/2.

In sum, the order flow at date 1, $\tilde{f}_1$, is either completely uninformative about the raw signal, s, or fully revealing. In the former case, the return from date 0 to date 1, denoted $r_1=(p_1-p_0)$, is zero. Otherwise this return is strictly positive if $s=1$ and strictly negative if $s=0$. Thus, the probability of a price movement at date 1 ($p_1 \neq p_0$) is given by the probability that the order flow is fully revealing, i.e., $\Pr(p_1 \neq p_0)=\min\{\alpha_1, 1\}$. This probability increases with the mass of speculators buying the raw signal, α_1, because, as their mass increases, their aggregate order size becomes larger relative to that of liquidity traders. Thus, speculators trading on the raw signal account for a larger fraction of the total order flow, which therefore becomes more informative. As a result, the price at date 1 becomes more responsive to trades at this date.

At $t=2$, speculators who have purchased the processed signal observe (s,u) and the price realized in period 1, p_1. Hence, we denote their trading strategy by $x_2(s,u,p_1)$ and their expected trading profit (per capita) is:

$$\pi_2(\alpha_1,\alpha_2,s,u,p_1)=x_2(s,u,p_1)(\mathsf{E}[V|s,u]-\mathsf{E}[p_2|s,u,p_1]).$$

In the rest of the paper, we denote by $\pi_2^c(\alpha_2)$ and $\pi_2^{nc}(\alpha_2)$, the expected profits of a speculator who buys the processed signal conditional on (i) a change ('c') in the price at date 1 (i.e., $p_1 \neq p_0$) and (ii) no change ('nc') in the price at date 1 (i.e., $p_1=p_0=1/2$), respectively.

Proposition 2. *The equilibrium at date $t=2$ is as follows:*

1. *If the processed signal is $(s,0)$, speculators who receive this signal buy one share if the price in the first period is smaller than $\frac{1}{2}$ (i.e., $x_2(s,0,p_1)=1$ if $p_1<1/2$); sell one share if the price in the first period is greater than $\frac{1}{2}$ (i.e., $x_2(s,0,p_1)=-1$ if $p_1>1/2$); and do not trade otherwise (i.e., $x_2(s,0,1/2)=0$). If instead the processed signal is $(s,1)$, speculators who receive this signal buy one share if*

$s = 1$ *(i.e.,* $x_2(1, 1, p_1) = 1$*) and sell one share if* $s = 0$ *(i.e.,* $x_2(0, 1, p_1) = -1$*). Speculators who do not receive the processed signal do not trade at date* 2*.*

2. *If* $p_1 = \mu(1) = \frac{1+\theta}{2}$ *then the asset price at date 2 is:*

$$p_2^*(f_2) = \begin{cases} \frac{1}{2} & \text{if } f_2 \in [f_2^{min}, -1 + \alpha_2], \\ \frac{1+\theta}{2} & \text{if } f_2 \in [-1 + \alpha_2, 1 - \alpha_2], \\ 1 & \text{if } f_2 \in [1 - \alpha_2, f_2^{max}]. \end{cases}$$

3. *If* $p_1 = \mu(0) = \frac{1-\theta}{2}$ *then the asset price at date 2 is:*

$$p_2^*(f_2) = \begin{cases} 0 & \text{if } f_2 \in [f_2^{min}, -1], \\ \frac{1-\theta}{2} & \text{if } f_2 \in [-1 + \alpha_2, 1 - \alpha_2], \\ \frac{1}{2} & \text{if } f_2 \in [1 - \alpha_2, f_2^{max}]. \end{cases}$$

4. *If* $p_1 = \frac{1}{2}$ *then the asset price at date 2 is:*

$$p_2^*(f_2) = \begin{cases} 0 & \text{if } f_2 \in [f_2^{min}, -1], \\ \frac{1-\theta}{2-\theta} & \text{if } f_2 \in [-1, \min\{-1 + \alpha_2, 1 - \alpha_2\}], \\ \frac{1}{2} & \text{if } f_2 \in [\min\{-1 + \alpha_2, 1 - \alpha_2\}], \max\{-1 + \alpha_2, 1 - \alpha_2\}] \\ \frac{1}{2-\theta} & \text{if } f_2 \in [\max\{-1 + \alpha_2, 1 - \alpha_2\}, 1] \\ 1 & \text{if } f_2 \in [1, f_2^{max}]. \end{cases}$$

5. *The ex-ante expected profit of speculators who buy the processed signal,* $\bar{\pi}_2(\alpha_1, \alpha_2) \overset{def}{=} \mathsf{E}[\pi_2(\alpha_1, \alpha_2, s, u, p_1)]$*, is:*

$$\bar{\pi}_2(\alpha_1, \alpha_2) = \alpha_1 \pi_2^c(\alpha_2) + (1 - \alpha_1)\pi_2^{nc}(\alpha_2), \tag{6}$$

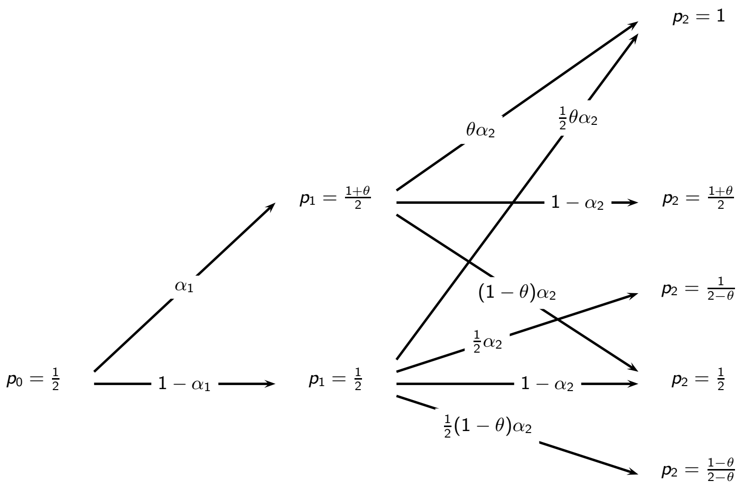
where $\pi_2^c(\alpha_2) = \max\{\theta(1 - \theta)(1 - \alpha_2), 0\}$ *and*

$$\pi_2^{nc}(\alpha_2) = \begin{cases} \frac{\theta}{2(2-\theta)}(2 - \theta - \alpha_2) & \text{if } \alpha_2 \leq 1 \\ \frac{\theta}{2}\frac{1-\theta}{2-\theta}(2 - \alpha_2) & \text{if } 1 < \alpha_2 \leq 1, \\ 0 \text{ if } & \alpha_2 > 2, \end{cases} \tag{7}$$

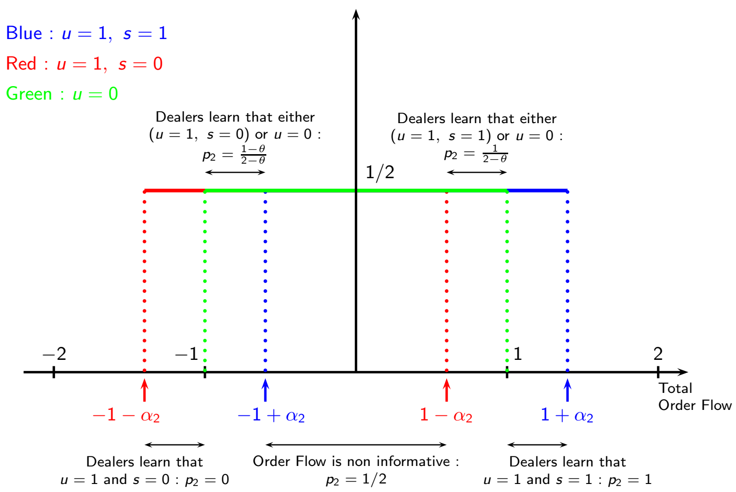
Figure 3: Equilibrium Price Dynamics

Panel A shows the possible equilibrium paths for the asset price when $s = 1$. Panel B shows the distribution of the order flow at date 2 for each possible realization of the processed signal when there is no change in the price at date 1 ($p_1 = p_0 = 1/2$).

(A) Equilibrium Price Dynamics when s=1



(B) Equilibrium Distribution of the Order Flow at Date $t = 2$ (f_2), when $p_1 = 1/2$, and $\alpha_2 < 1$



The trading decision of speculators who receive the processed signal at date 2 depends on whether $u = 1$ or $u = 0$. When $u = 1$, the processed signal confirms the raw signal s. Thus, speculators trade on the processed signal as they trade on the raw signal, i.e., they buy the asset if $s = 1$ (the asset payoff is high) and sell it if $s = 0$ (the asset payoff is zero). Hence, *conditional* on $u = 1$, speculators' trading decision at date 2 is independent from the price of the asset at the end of the first period.

In contrast, when $u = 0$, the processed signal invalidates the raw signal and speculators receiving the processed signal expect the payoff of the asset to be $1/2$. Their trading decision is then determined by the latest price of the asset, i.e., p_1. If $p_1 > 1/2$, they optimally sell the asset because they expect that, on average, their sell orders will execute at a price greater than their valuation for the asset ($1/2$). Symmetrically, if $p_1 < 1/2$, they optimally buy the asset. Finally, if $p_1 = 1/2$ and $u = 0$, not trading is weakly dominant for speculators who receive the processed signal because they expect their order to execute at a price equal to their valuation for the asset, i.e., $1/2$.[17]

Panel A of Figure 3 shows the possible equilibrium price paths when $s = 1$ (the case in which $s = 0$ is symmetric) and the transition probabilities from the price obtained at date 1 to the price at date 2 (when $\alpha_1 \leq 1$ and $\alpha_2 \leq 1$).[18]

When $s = 1$, speculators who receive the raw signal buy the asset at date 1 and, with probability α_1, the market maker infers from the order flow that $s = 1$ and sets a price equal to $p_1 = \mu(1) = \frac{1+\theta}{2} > p_0$. In this case, after trading at date 1, the only remaining source of uncertainty for the market maker is about u. At date 2, with probability θ, the processed signal confirms the raw signal (i.e., $(s, u) = (1, 1)$). Hence, speculators who receive this signal also buy the asset and, with probability α_2, their demand is so strong that the market maker infers that $V = 1$. In this case, the price goes up at date 2 relative to the price at date 1. The overall unconditional probability of two consecutive up movements in the price is therefore $(\theta\alpha_1\alpha_2)/2$.[19] Alternatively, with probability $(1 - \theta)$, the processed signal invalidates the raw signal (i.e., $(s, u) = (1, 0)$). Hence, speculators sell the asset in period 2 because, given their information, it is overpriced. In this case, with probability α_2, their supply is strong enough to push the price back to its initial level and they in fact correct the noise injected by speculators at date 1 into prices. Thus, the unconditional probability of an up price movement followed by a down movement is $((1-\theta)\alpha_1\alpha_2)/2$. Finally, in either case, there is a probability $(1-\alpha_2)$ that the order flow at date 2 is uninformative. In this case, the price at date 2 is equal to the price at date 1.

When the market maker does not infer the raw signal, s, from trades at date 1, his

[17] The reason is that, in this case, speculators expect (i) liquidity traders' aggregate demand for the asset to be zero on average and (ii) other speculators' demand for the asset to be zero as well. Hence, a speculator expects the price at date 2 to be identical to the price at date 1 because his demand is negligible compared to speculators' aggregate demand.

[18] Transition probabilities are different when $\alpha_2 > 1$.

[19] The unconditional probability of a given price path in equilibrium is obtained by multiplying the conditional likelihood of this path by $1/2$ because $s = 1$ or $s = 0$ with equal probabilities.

inference problem at date 2 is more complex since he then knows neither s, nor u. This explains why there are more possible realizations for the equilibrium price at date 2 when there is no price change at date 1. For instance, suppose that the market observes a realization of the aggregate order flow at date 2 in the interval $[-1, -1+\alpha_2]$. As Panel B of Figure 3 shows, this realization is consistent with three possible realizations of the processed signal $(1,0)$, $(0,0)$, or $(0,1)$. Thus, the market maker sets a price equal to $p_2 = \mathsf{E}(v \mid (s,u) \in \{(1,0),(0,0),(0,1)\}) = (1-\theta)/(2-\theta) < 1/2$. This explains why, even though $s = 1$, the price might decrease from date 1 to date 2 when it has not changed at date 1.

The expected profit from trading on a given signal (raw or processed) decreases with the number of speculators buying this signal (that is, $\frac{\partial \bar{\pi}_1(\alpha_1)}{\partial \alpha_1} \leq 0$ and $\frac{\partial \bar{\pi}_2(\alpha_1,\alpha_2)}{\partial \alpha_2} \leq 0$). Indeed, as more speculators trade on a signal, the order flow (or price) becomes more informative about this signal and, as a result, expected profit from trading on this signal drops. For instance, when α_1 increases, the expected profit of trading on the raw signal declines because, as explained previously, the likelihood that the order flow reveals speculators' signal at date 1 becomes higher. This effect is standard in models of informed trading (e.g., Grossman and Stiglitz (1980) or Kyle (1985)).[20]

More surprisingly, the next corollary shows that investors trading on the processed signal can in fact benefit from a more informative price at date 1. That is, for some parameter values, their expected profit is higher when the market maker learns the raw signal at date 1 (and adjusts his price accordingly) than when he does not. Let denote $\hat{\alpha}_2(\theta) = \frac{(1-2\theta)(2-\theta)}{2(2-3\theta+\theta^2)-1}$. Observe that $\hat{\alpha}_2(\theta) > 0$ iff $\theta < 1/2$ and that $\hat{\alpha}_2(\theta)$ goes to $2/3$ as θ goes to zero.

Corollary 1. *The expected profit from trading on the processed signal is larger when the market maker learns the raw signal (the order flow is fully revealing) at date* 1 *than when he does not (i.e.,* $\pi_2^c(\alpha_2) > \pi_2^{nc}(\alpha_2)$*) when* $\alpha_2 < \hat{\alpha}_2(\theta)$ *and* $\theta \leq 1/2$*. Otherwise it is smaller.*

The intuition for this finding is as follows. Suppose that the price reflects the raw signal, s, at the end of period 1. If this signal is valid then speculators receiving the

[20]When $\alpha_1 \geq 1$, the expected profit from trading on the raw signal, s, is nil because the mass of speculators trading on signal is so large relative to the mass of liquidity traders that the order flow at date 1 is always fully revealing (the interval $[-1+\alpha_1, 1-\alpha_1]$ is empty). For a similar reason, the expected profit from trading on the processed signal, (s,u), is zero when the mass of speculators trading on the processed signal is twice the mass of liquidity traders, i.e., when $\alpha_2 \geq 2$. The trading strategy that exploits the processed signal has a larger "capacity" (break even for a larger number of speculators) because it is more difficult for market makers to infer information about the processed signal from the order flow.

processed signal obtain a smaller expected profit than if the price had not changed since the asset price already impounds part of their information about the asset payoff. This is a standard logic in models of information acquisition. However, the logic is reversed if the raw signal is noise. Indeed, if the price at date 1 reflects the raw signal, speculators receiving the processed signal can make a profit by correcting the noise in the price, either by selling the asset if the price increased in the last period or buying it if the price decreased. This profit opportunity does not exist if the price has not changed at date 1. For this reason, if the raw signal is noise, speculators receiving the processed signal are better off when the price reflects the raw signal at date 1 than when it does not. This second effect dominates if the likelihood that the signal is noise is large enough ($\theta \leq 1/2$) and competition among speculators receiving the processed signal is not too intense ($\alpha_2 < \hat{\alpha}_2(\theta)$). In this case, on average, speculators obtain a larger profit when the first period price reflects the raw signal than when it does not.

The previous result implies that an increase in the demand for the raw signal can have a positive effect on the expected profit of speculators who received the processed signal. This effect again is non standard. Indeed, in standard models of trading with asymmetric information, the expected profit of informed investors usually decrease with the mass of informed investors. In contrast, in our setting, an increase in the mass of speculators informed about the raw signal can in fact result in larger expected profits for speculators who receive the processed signal.

To see this, observe that the marginal effect of an increase in the demand for the raw signal on the unconditional expected profit of trading on the processed signal (given by eq.(6)) is:

$$\frac{\partial \bar{\pi}_2(\alpha_1, \alpha_2)}{\partial \alpha_1} = \pi_2^c(\alpha_2) - \pi_2^{nc}(\alpha_2). \tag{8}$$

Thus, if $\pi_2^c(\alpha_2) > \pi_2^{nc}(\alpha_2)$, an increase in the demand for the raw signal (α_1) increases the unconditional expected profit of trading on the processed signal. Intuitively, it raises the likelihood that the price will reflect the raw signal in the first period. This is beneficial for speculators who trade on the processed signal if their expected profit is higher when the first period price reflects the raw signal, that is, if $\alpha_2 < \hat{\alpha}_2(\theta)$ and $\theta \leq 1/2$. The next corollary follows.

Corollary 2. *The expected profit from trading on the processed signal, $\bar{\pi}_2(\alpha_1, \alpha_2)$, increases with the demand for the raw signal, α_1, if and only if $\alpha_2 < \hat{\alpha}_2(\theta)$ and $\theta \leq 1/2$.*

In sum, an increase in the demand for the raw signal (α_1) can either strengthen or lower the value of the processed signal (i.e., the expected profit from trading on this signal). Thus, an increase in the equilibrium demand for the raw signal could either increase or reduce the demand for the processed signal (see Figure 4). To study this issue, we analyze the equilibrium of the market for information (the prices and demands for the raw and the processed signals at date 0) in the next section.

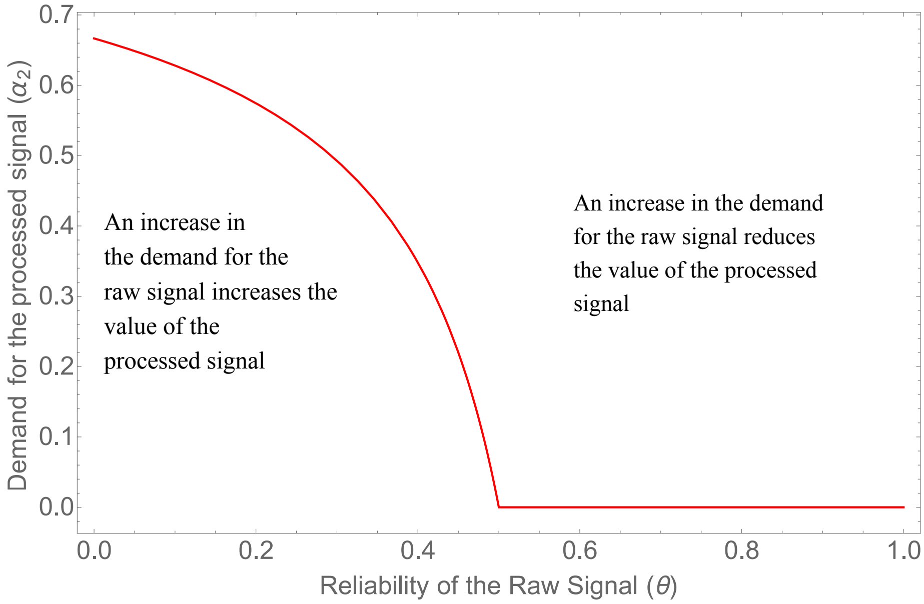


Figure 4: This figure shows the sets of values of θ and α_2 for which a marginal increase in the demand for the raw signal (α_1) increases or decreases the value of the processed signal for speculators. The red curve is the threshold $\hat{\alpha}_2(\theta)$ defined in Corollary 2.

5 Equilibrium in the Market for Information

In this section, we derive the fees charged by information sellers and the resulting equilibrium demands (α_2^e and α_1^e) for each type of signal. Producing information goods involves large fixed costs but negligible marginal costs (see, for instance, Shapiro and Varian (1999) or Veldkamp (2011), Chapter 8 and references therein). For instance, Shapiro and Varian (1999) write (on page 21): "*Information is costly to produce but cheap to reproduce [...]. This cost structure leads to substantial economies of scale.*" Thus, as in Veldkamp (2006a,b), we assume that information sellers bear a fixed cost to produce their signal (denoted C_p for the seller of the processed signal and C_r for seller of the raw signal) and zero cost to distribute it. For instance, C_r represents the cost of collecting data and designing an algorithm to extract the raw signal s from these data. This cost is independent from the number of speculators buying the raw signal and the marginal cost of distributing this signal to an extra buyer is zero.[21]

As in Veldkamp (2006a,b), we also assume that markets for information are competitive and perfectly contestable. This means that, in equilibrium, the buyers and the seller of a given signal make zero expected profits. We first derive the equilibrium in the market for the processed signal, holding the demand for the raw signal (α_1) fixed. This is without loss of generality because the equilibrium value of α_1 is independent of the equilibrium value of α_2 (while the reverse is not true; see below). Thus, α_1 can be treated as a parameter in the analysis of the equilibrium of the market for the processed signal.

The demand for a signal is capped by the (relative) mass of speculators, $\bar{\alpha}$. As shown below, this upper bound on the demand for a given signal is never binding when $\bar{\alpha} \geq 2$. To simplify the exposition, we assume that this is the case. In this way, we eliminate corner cases in which the demand for a given signal hits the upper bound $\bar{\alpha}$ and becomes therefore insensitive to a change in parameters (e.g., C_r). This reduces the number of cases to discuss when presenting the equilibrium of the market for each signal. For completeness, we show in Section 2 of the on-line appendix that the results are unchanged when $\bar{\alpha} < 2$.

From the viewpoint of each speculator, the cost of acquiring the processed signal is the fee, F_p, charged by seller of this signal. Each speculator takes this fee as given. Let $\bar{\pi}_2^{net}(\alpha_1, \alpha_2, F_p) \stackrel{\text{def}}{=} \bar{\pi}_2(\alpha_1, \alpha_2) - F_p$ be the expected profit from trading on the processed signal in equilibrium net of the fee paid to obtain this signal. Moreover, let $\bar{\Pi}_2^{seller}(\alpha_2, F_p) \stackrel{\text{def}}{=} \alpha_2 \times F_p - C_p$ be the expected profit of the seller of the processed signal. Finally let (F_p^e, α_2^e) be the equilibrium fee and the equilibrium demand for the processed signal. We say that the market for the processed signal is *active* if $\alpha_2^e > 0$.

If the market for the processed signal is active, in a competitive equilibrium, the demand for the processed signal and the fee charged for this signal must be such that the speculators buying the processed signal and the seller of this signal just break even. That is, when $\alpha_2^e > 0$, (F_p^e, α_2^e) solve:

$$\text{Zero profit for speculators:} \quad \bar{\pi}_2^{net}(\alpha_1, \alpha_2^e, F_p^e) = \bar{\pi}_2(\alpha_1, \alpha_2^e) - F_p^e = 0. \tag{9}$$

[21] Consider a firm like iSentium (see http://www.iSENTIUM.com/) that specializes in selling investment signals extracted from social medias, like tweeter. For this firm, the total cost of producing a signal will comprise the cost of subscribing to the complete Twitter stream (\$30,000 a month; see "How investors are using social medias to make money", Fortune, December 2015)) and developing algorithms to extract signals from this stream. These costs are fixed, i.e., they do not depend on the number of investors buying iSentium's signals. They correspond to C_r in our model. iSentium charges a fee of \$15,000 per month to each subscriber buying its signals. This corresponds to the fee F_r in our model.

and

$$\text{Zero profit for the information seller:} \quad \bar{\Pi}_2^{seller}(\alpha_2^e, F_p^e) = \alpha_2^e \times F_p^e - C_p = 0.^{22} \tag{10}$$

Condition (9) implies that, at equilibrium, a speculator is indifferent between buying the processed signal or not trading at date 2 (taking other speculator's decisions and the fee for the processed signal as given). If this was not the case then α_2^e would not be the equilibrium demand for the processed signal since either additional speculators would benefit from buying the signal or some buyers of the signal would be better off not buying it. Condition (10) is necessary to preclude profitable entry by another seller of the processed signal and sufficient if F_p^e is the smallest possible fee among all possible equilibrium fees. Thus, when there are multiple solutions (F_p^e, α_2^e) to eq.(9) and (10), we select the one with the smallest fee, since other fees could profitably be undercut by another information seller.

When the market for the processed signal is active ($\alpha_2^e > 0$), Condition (9) implies that the *aggregate* net expected profit (denoted $\bar{\pi}_2^{net,a}(\alpha_1, \alpha_2^e)$) of speculators buying the processed signal is zero. Thus, using eq.(10), we deduce that when $\alpha_2^e > 0$ then:

$$\bar{\pi}_2^{net,a}(\alpha_1, \alpha_2^e) = \alpha_2^e \pi_2^{net}(\alpha_1, \alpha_2^e, F_p^e) = \pi_2^{gross,a}(\alpha_1, \alpha_2^e) - C_p = 0, \tag{11}$$

where $\pi_2^{gross,a}(\alpha_1, \alpha_2) \stackrel{\text{def}}{=} \alpha_2 \bar{\pi}_2(\alpha_1, \alpha_2)$ denotes the aggregate *gross* expected profit for speculators trading on the processed signal, for given values of α_1 and α_2. Condition (11) is equivalent to:

$$\pi_2^{gross,a}(\alpha_1, \alpha_2^e) = C_p. \tag{12}$$

Thus, when the market for the processed signal is active, the equilibrium demand for this signal is such that the aggregate *gross* expected profit of speculators buying this signal is equal to its production cost.

[22] Information sellers have rational expectations about the demand for the signal they sell. That is, they expect this demand to be equal to the actual equilibrium demand, α_2^e.

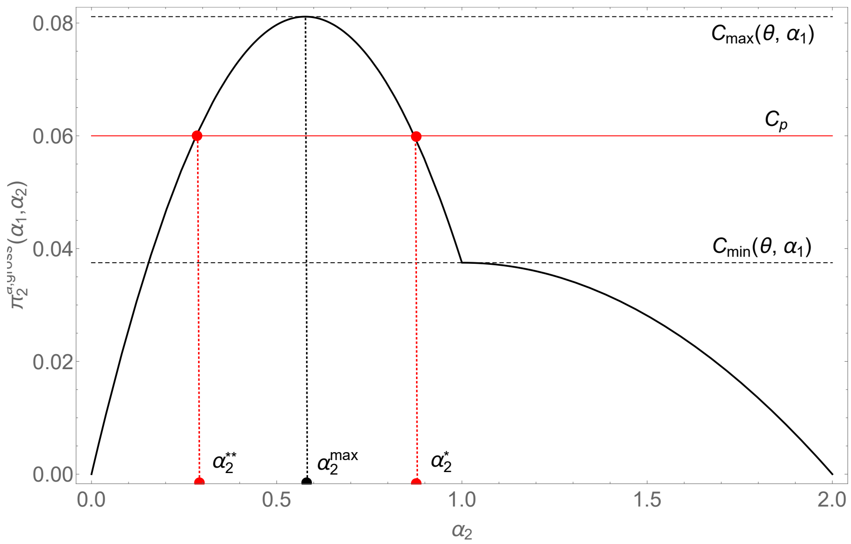


Figure 5: This figure represents speculators' aggregate gross expected profit from trading on the processed signal as a function of the demand for this signal, α_2.

Speculators' aggregate gross expected profit, $\pi_2^{gross,a}(\alpha_1, \alpha_2)$, is hump-shaped in the demand for the processed signal, α_2, holding α_1 fixed (see Figure 5). We denote by $\alpha_2^{max}(\alpha_1, \theta)$ the demand for the processed signal that maximizes the aggregate gross expected trading profit from trading on this signal. Using eq.(6), we obtain:

$$\alpha_2^{max}(\alpha_1, \theta) = \frac{(2-\theta)(1-(2\theta-1)\alpha_1)}{2(1+(2(2-\theta)(1-\theta)-1)\,\alpha_1)} \leq 1. \tag{13}$$

We deduce from eq.(6) that the maximum aggregate gross expected trading profit from trading on the processed signal, denoted $C_{max}(\theta, \alpha_1)$, is:

$$C_{max}(\theta, \alpha_1) \stackrel{\text{def}}{=} \pi_2^{gross,a}(\alpha_1, \alpha_2^{max}) = \frac{\theta(1-(2\theta-1)\alpha_1)\alpha_2^{max}}{4}. \tag{14}$$

First, consider the case in which $C_p < C_{max}(\theta, \alpha_1)$, as assumed in Figure 5. For $\alpha_2 \in [\alpha_2^{max}, 2]$, speculators' aggregate gross expected profit decreases in α_2 from $C_{max}(\theta, \alpha_1)$ to zero. Thus, there is a unique $\alpha_2^* \in (\alpha_2^{max}, 2)$ solving eq.(12) for $0 < C_p < C_{max}$. In general, as Figure 5 shows, there is another value of α_2, denoted α_2^{**}, solving eq.(12). This value is necessarily on the increasing segment of speculators' aggregate gross expected profit (see Figure 5) since α_2^* is the unique solution on the decreasing segment, as we just explained. Thus, $\alpha_2^{**} < \alpha_2^{max} < \alpha_2^* < 2$.

Therefore, either $\alpha_2^e = \alpha_2^*$ or $\alpha_2^e = \alpha_2^{**}$ when the market for processed information is active. In the former case, the zero expected profit condition for the information seller imposes $F_2^e = C_p/\alpha_2^*$ while in the latter it imposes $F_2^e = C_p/\alpha_2^{**}$. Hence, the information seller's fee is smaller in the first case since $\alpha_2^{**} < \alpha_2^*$. Thus, when the market for the processed signal is active, the unique competitive equilibrium of this market is $(\alpha_2^e, F_2^e) = (\alpha_2^*, C_p/\alpha_2^*)$. As $\alpha_2^* < 2 \leq \bar{\alpha}$, the constraint $\alpha_2^e < \bar{\alpha}$ is never binding.

Now consider the case in which $C_p \geq C_{max}(\theta, \alpha_1)$. In this case, eq.(12) has no solution because, for any α_2, the gross aggregate profit from trading on the processed signal is smaller than C_p (see Figure 5). Thus, there is no fee for the processed signal at which transactions between the buyers and the seller of the processed signal are mutually profitable. Consequently, when $C_p \geq C_{max}(\theta, \alpha_1)$, the market for the processed signal is inactive, i.e., $\alpha_2^e = 0$.

The next lemma summarizes the previous discussion by providing the closed form solution for the equilibrium demand for the processed signal, α_2^e, and the corresponding fee charged by the seller of this signal.

Lemma 1. *Let* $C_{min}(\theta, \alpha_1) = \frac{\theta(1-\theta)(1-\alpha_1)}{2(2-\theta)}$. *The competitive equilibrium of the market for the processed signal is unique.*

1. *If* $C_p < C_{max}(\theta, \alpha_1)$, *the equilibrium demand for the processed signal is:*

$$\alpha_2^e(\theta, \alpha_1, C_p) = \begin{cases} \alpha_2^{max}(\theta, \alpha_1)\left(1 + \left(1 - \frac{C_p}{C_{max}(\theta,\alpha_1)}\right)^{\frac{1}{2}}\right) & \text{if } C_{min}(\theta, \alpha_1) \leq C_p \leq C_{max}(\theta, \alpha_1), \\ 1 + \left(1 - \frac{C_p}{C_{min}(\theta,\alpha_1)}\right)^{\frac{1}{2}} & \text{if } 0 \leq C_p < C_{min}(\theta, \alpha_1), \end{cases}$$

   and the equilibrium fee for the processed signal is $F_p^e = \frac{C_p}{\alpha_2^e}$.

2. *If* $C_p > C_{max}(\theta, \alpha_1)$, *there is no fee at which the seller and the buyers of the processed signal can trade in a mutually beneficial way. Thus, the processed signal is not produced in equilibrium and therefore* $\alpha_2^e = 0$.

Not surprisingly, as the fixed cost of producing the processed signal declines (starting from C_{max}), the fee charged by the seller of the processed signal falls and, therefore, the mass of speculators buying this signal increases ($\frac{\partial \alpha_2^e}{\partial C_p} \geq 0$; see Figure 5).

We derive the equilibrium of the market for the raw signal in a similar way (see Section 1 of the on-line appendix). We obtain the following.

Lemma 2. *The competitive equilibrium of the market for the raw signal is unique.*

1. *If* $C_r < \frac{\theta}{8}$, *the equilibrium demand for the raw signal is:*

$$\alpha_1^e(\theta, C_r) = \frac{1}{2} + \left(\frac{1}{4} - \frac{2C_r}{\theta}\right)^{\frac{1}{2}} \tag{15}$$

and the equilibrium fee for the raw signal is $F_r^e = \frac{C_r}{\alpha_1^e}$.

2. *If $C_r \geq \frac{\theta}{8}$, there is no fee at which the seller and the buyers of the raw signal can trade in a mutually beneficial way. Thus, the raw signal is not produced in equilibrium and therefore $\alpha_1^e = 0$.*

Thus, in equilibrium, the demand for the raw signal, α_1^e, increases when the cost of producing this signal decreases.[23] Through this channel, a decrease in the cost of producing the raw signal, C_r, has also an effect on the equilibrium demand for the processed signal since the latter is influenced by the demand for the raw signal (see Corollary 2). The next proposition analyzes this effect. Let $\bar{C}_r(\theta) = \frac{\theta}{2}\left(\frac{1}{4} - \max\left(\frac{(1-\theta)^2+\theta^2}{(1-2\theta)[2(1-\theta)(2-\theta)-1]} - \frac{1}{2}, 0\right)^2\right)$ and $\bar{C}_p(\theta) = \frac{\theta(1-\theta)^2(2-\theta)(1-2\theta)}{(2(1-\theta)(2-\theta)-1)^2}$.

Proposition 3. 1. *For $\theta > \frac{\sqrt{2}-1}{\sqrt{2}}$, a decrease in the cost of producing the raw signal reduces the equilibrium demand for the processed signal ($\frac{\partial \alpha_2^e}{\partial C_r} > 0$).*

2. *For $\theta \leq \frac{\sqrt{2}-1}{\sqrt{2}}$, a decrease in the cost of the raw signal increases the equilibrium demand for the processed signal (i.e., $\frac{\partial \alpha_2^e}{\partial C_r} < 0$) if $C_r < \bar{C}_r(\theta)$ and $C_p > \bar{C}_p(\theta)$. Otherwise, a decrease in the cost of producing the raw signal reduces the equilibrium demand for the processed signal.*

As shown in Corollary 2, an increase in the demand for the raw signal can either reduce or increase the gross expected profit from trading on the processed signal. It increases this profit if and only if $\alpha_2^e < \hat{\alpha}_2(\theta)$ (see Corollary 2). In the proof of Proposition 3, we show that this condition is equivalent to $\theta \leq \frac{\sqrt{2}-1}{\sqrt{2}}$, $C_r < \bar{C}_r(\theta)$, and $C_p > \bar{C}_p(\theta)$. In this case, a decrease in the cost of producing the raw signal triggers, directly, an increase in the equilibrium demand for the raw signal and thereby, *indirectly*, an increase in the expected profit from trading on the processed signal, holding the demand for this signal constant. As a result, the demand for the processed signal increases.

Otherwise (e.g., when $\theta > \frac{\sqrt{2}-1}{\sqrt{2}}$), an increase in the demand for the raw signal undermines the expected gross profit from trading on the processed signal. Consequently, in this case, the demand for the processed signal declines when the cost of producing the raw signal decreases, as shown in Figure 6.

[23] The largest possible demand for the raw signal is obtained when $C_r = 0$ and is equal to 1. Thus, the constraint $\alpha_1^e < \bar{\alpha}$ is never binding since $\bar{\alpha} \geq 2$.

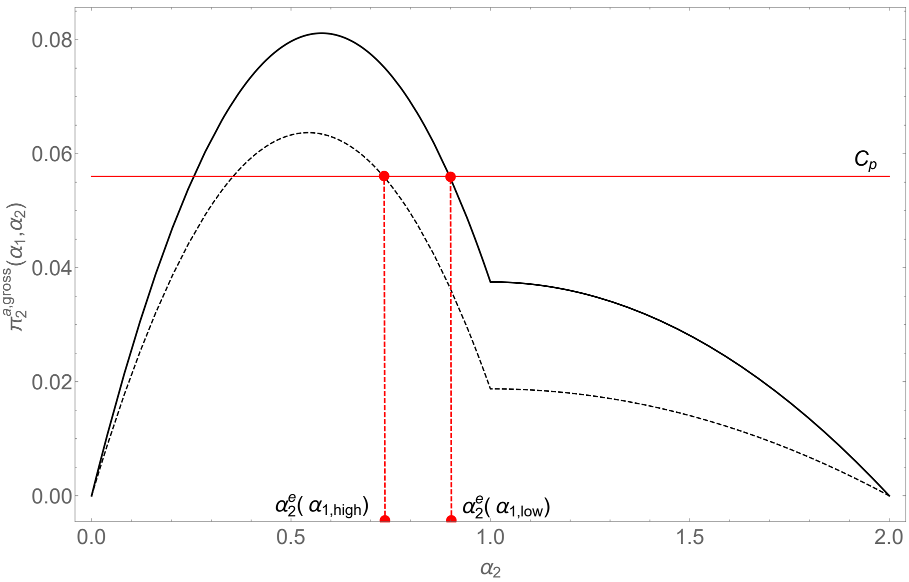


Figure 6: This figure represents speculators' aggregate profit from trading on the processed signal as a function of the demand for this signal for two different levels of the demand for the raw signal: (i) High ($\alpha_{1,high}$) and (ii) Low ($\alpha_{1,low}$). The former case is obtained when the cost of producing the raw signal is lower than in the latter case. The corresponding equilibrium demands for the processed signal in each case are, respectively, $\alpha_2^e(\alpha_{1,high})$ and $\alpha_2^e(\alpha_{1,low})$.

As the next proposition shows, this crowding out of "good" signals by "bad" signals can lead to a complete disappearance of the market for the processed signal (a discontinuous drop to zero of the demand for this signal), even though this market would be viable in the absence of a market for the raw signal.

Proposition 4. *Suppose $\theta > \frac{\sqrt{2}-1}{\sqrt{2}}$ and $\frac{\theta(1-\theta)}{4} < C_p < \frac{\theta(2-\theta)}{8}$. There exists a threshold $\widehat{C}_r(\theta, C_p)$ (defined in the proof of the proposition) such that if $C_r \geq \widehat{C}_r$, $\alpha_2^e \geq \alpha_2^{max} > 0$ while if $C_r < \widehat{C}_r$, $\alpha_2^e = 0$.*

Thus, when $\theta > \frac{\sqrt{2}-1}{\sqrt{2}}$, the demand for the processed signal discontinuously drops to zero when C_r passes below $\widehat{C}_r(\theta, C_p)$. Indeed, in this case, as the cost of producing the raw signal declines, more speculators choose to buy this signal, which reduces the expected return from trading on the processed signal. If the cost of producing the raw signal is just equal to the threshold $\widehat{C}_r$, the largest possible value for the gross expected trading profit of speculators trading on the processed signal is just equal to the cost of producing this signal, C_p. At this point, any further decrease in the production cost of the raw signal implies that the aggregate gross expected trading profit for speculators trading on the processed signal is smaller than the cost of producing this signal. Thus, there is no fee at which the sellers and buyers of the processed signal can find profitable to trade together.

Hence, the market for the process signal is not viable when $C_r < \widehat{C}_r(\theta, C_p)$ and therefore ceases to exist.

Figure 7 illustrates this result. As the cost of producing the raw signal, C_r declines, the demand for this signal increases in equilibrium (dotted line) while the demand for the processed signal declines (plain line). At $C_r = \hat{C}_r \approx 0.06$, the demand for the processed signal discontinuously drops from $\alpha_2^e \approx 0.6$ to zero.

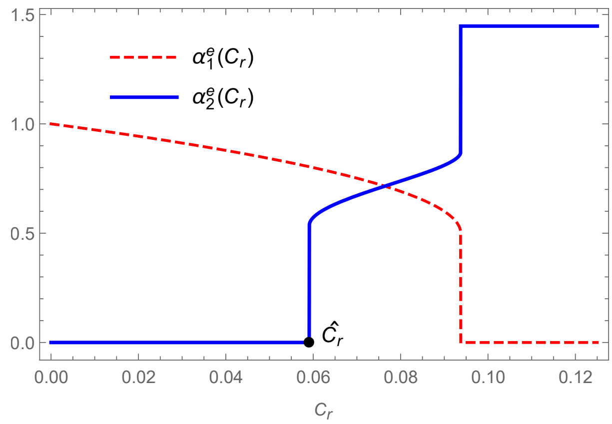


Figure 7: Equilibrium demands for the raw signal (red dotted line) and the processed signal (blue thick line) as a function of the cost of producing the raw signal, C_r (X-axis). Other parameters are $\theta = 0.75$ and $C_p = 0.06$.

6 Implications

6.1 Price Informativeness

We now study how a change in the cost of producing the raw signal affects price informativeness. In the absence of informed trading at dates 1 and 2 ($\alpha_1 = \alpha_2 = 0$), the asset price at each date is constant ($p_0 = p_1 = p_2 = 1/2$) and is therefore completely uninformative about the asset payoff. In this benchmark case, the average squared pricing error (the difference between the asset payoff and its price) is therefore $\mathsf{E}[(\tilde{V} - p_0)^2] = 1/4$ at dates 1 and 2. We measure price informativeness at date t, denoted $\mathcal{E}_t(C_r, C_p)$, by the *difference* between the average pricing error in the benchmark case (completely uninformative prices) and the average pricing error at date t in equilibrium, i.e., by:

$$\mathcal{E}_t(C_r, C_p) = \frac{1}{4} - \mathsf{E}[(\tilde{V} - p_t^*)^2] \tag{16}$$

The more informative is the price at date t in equilibrium, the higher is $\mathcal{E}_t(C_r, C_p)$. The *highest* possible value for $\mathcal{E}_t(C_r, C_p)$ is obtained if the price at date t is fully informative

$(p_t = V)$ and is therefore equal to $1/4$. The smallest possible value is equal to zero and is obtained when the price at date t is uninformative.[24] Thus, $\mathcal{E}_t(C_r, C_p)$ belongs to $[0, 1/4]$.

We refer to the informativeness of the price in the first period, $\mathcal{E}_1(C_r, C_p)$, as "short run price informativeness" and to the informativeness of the price in the second period, $\mathcal{E}_2(C_r, C_p)$, as "long run price informativeness" (the notion of short and long run is relative to the moment at which new data for assessing the asset payoff becomes available, i.e., date 1).

The next corollary studies how a change in the cost of producing the processed signal (C_p) affects price informativeness in equilibrium (i.e., accounting for the effects of a change in this cost on equilibrium fees and demands for the processed and the raw signals).

Corollary 3. *A reduction in the cost of producing the processed signal has no effect on short run price informativeness ($\frac{\partial \mathcal{E}_1(C_r,C_p)}{\partial C_p} = 0$) while it (weakly) increases long run price informativeness ($\frac{\partial \mathcal{E}_2(C_r,C_p)}{\partial C_p} \leq 0$).*

A decrease in the cost of producing the processed signal raises the demand for the processed signal in equilibrium and therefore leads to more informative prices at date 2. This effect is standard in models with endogenous information acquisition (e.g., Grossman and Stiglitz (1980)): when the cost of producing information declines, the demand for information increases and prices become more informative.

Our main new result regarding price informativeness is that this logic does not necessarily apply when one considers a decline in the cost of producing the raw signal, C_r, i.e., signals that can be produced and delivered *before* the, more precise, processed signal is delivered. Indeed, even though a decline in this cost improves price informativeness in the short run, it can *impair* long run price informativeness, as shown by the next proposition.

Proposition 5. *A reduction in the cost of producing the raw signal (weakly) increases short run price informativeness ($\frac{\partial \mathcal{E}_1(C_r,C_p)}{\partial C_r} \leq 0$). However, if $C_p \leq C_{min}(\theta, \alpha_1^e)$ then it reduces long run price informativeness.*

Short run price informativeness increases when the cost of producing the raw signal declines because it leads more speculators to buy this signal. As the raw signal is sometimes truly informative ($\theta > 0$, as otherwise no investor buys the raw signal), the increase in the mass of speculators trading on the raw signal makes the asset price more

[24] As $p_t = \mathsf{E}[V|\Omega_t]$, we have $\mathsf{E}[(\tilde{V} - p_t^*)^2] = \mathsf{E}[Var[V|\Omega_t]]$. Thus, $\mathcal{E}_t(C_r, C_p) = Var(V) - E[Var[V|\Omega_t]]$. Hence, price informativeness at date t is higher when, on average, the price at this date provides a more accurate estimate of the asset payoff.

informative at date 1. However, when $C_p \leq C_{min}(\theta, \alpha_1^e)$, this effect triggers a drop in the demand for the processed signal because it reduces the expected profit from trading on this signal (see Corollary 3 and Figure 7). This indirect effect of a reduction in the cost of producing the raw signal tends to decrease long run price informativeness.

In Section 5 of the on-line appendix, we show that this result can hold even when $C_{min}(\theta, \alpha_1^e) < C_p < C_{max}(\theta, \alpha_1)$ (that is, the condition on C_p in Proposition 5 is sufficient but not necessary). However, in this case, there are also parameter values for which a reduction in the cost of producing the raw signal improves long run price informativeness because it raises the expected trading profit from trading on the processed signal and thereby the demand for this signal (see Corollary 3). Last, when $C_p > C_{max}(\theta, \alpha_1)$, the market for the processed signal is inactive. Thus, information production stops after date 1. In this case, the informativeness of the price at date 2 is equal to that at date 1 and therefore it increases when the cost of producing the raw signal, C_r, declines.

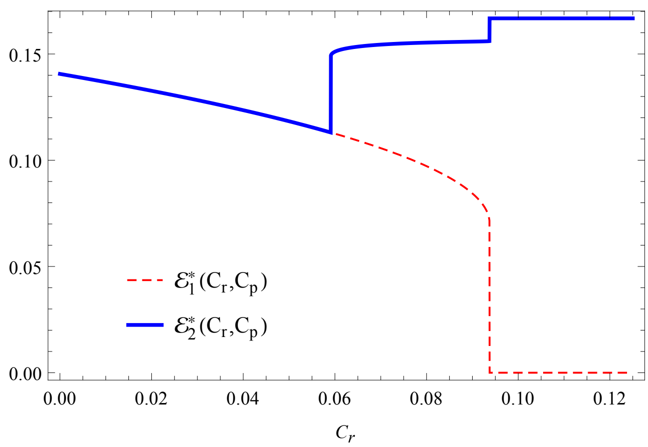


Figure 8: Price informativeness in the short run (red dotted line) and the long run (blue thick line) as a function of the cost of producing the raw signal C_r (X-axis). Other parameters are $\theta = 0.75$ and $C_p = 0.06$

Figure 8 illustrates these effects for specific parameter values. When the cost of producing the raw signal, C_r, is large (0.095 for the parameter values used in Figure 8), no investor buys this signal ($\alpha_1^e = 0$) and the demand for the processed signal is relatively high. As the cost of producing the raw signal declines, the demand for this signal starts increasing and consequently the demand for the processed signal, α_2^e, decreases (see Figure 7). Short run price informativeness increases but long run price informativeness declines (see Figure 8).[25] When $C_r = \widehat{C}_r \simeq 0.06$, the gross aggregate expected profit from trading

[25]Long run price informativeness is at least equal to short run price informativeness because the market maker has at least as much information at date 2 than he has at date 1 ($\Omega_1 \subset \Omega_2$). It is strictly higher (i.e., $\mathcal{E}_2(C_r, C_p) > \mathcal{E}_1(C_r, C_p)$) when $\alpha_2^e > 0$ because trades at date 2 contain new information

on the processed signal is just equal to the cost of producing this signal. At this point, if C_r decreases further, the demand for the processed signal discontinuously drops to zero (as implied by Corollary 4) and long run price informativeness drops discontinuously as well and becomes just equal to short run price informativeness. As C_r keeps declining, the demand for the raw signal increases. Hence, short run price informativeness improves and long run price informativeness does as well. Indeed, short run and long run price informativeness are now equal because there is no further investment in discovering the payoff of the asset after date 1.

Interestingly, even if $C_r = 0$ (i.e., $\alpha_1^e = 1$), price informativeness at date 2 is smaller than when the cost of producing the raw signal is so high ($C_r \geq \frac{\theta}{8}$) that there is no demand for the raw signal ($\alpha_1^e = 0$). For instance, for the parameter values considered in Figure 8, $\mathcal{E}_2(0, C_p) = 0.14$ while $\mathcal{E}_2(C_r, C_p) = 0.17$ for $C_r \geq \frac{\theta}{8}$. The next proposition shows that this conclusion holds more generally.

Proposition 6. *When $0 < C_p \leq \hat{C}_p(\theta)$, long run price informativeness is always smaller when the raw signal is free ($C_r = 0$) than when it is so costly that no investor buys it in equilibrium ($C_r > \frac{\theta}{8}$), where $\hat{C}_p(\theta)$ is defined in the appendix.*

Arguably, progress in information technologies have reduced *both* production costs for both low and high precision signals, i.e., both C_r and C_p in our model. However, as Proposition 6 shows, this evolution does not imply that long run price informativeness should improve. Indeed, for any level of the cost of producing the processed signal, if $\theta < 1$, there is always a sufficiently low value of the cost of producing the raw signal such that long run price informativeness is smaller than if there were no trading on the raw signal.[26]

The arrival of public news in financial markets (e.g., earnings announcements) offer trading opportunities for speculators because news often need to be interpreted and processed (see, for instance, Engelberg et al.(2012) for supporting evidence). One way to test our predictions is therefore to consider the evolution price informativeness after news arrival for a firm.

For instance, suppose public news arrives about the asset just before date 1. The prior distribution of V represents market participants' beliefs about the payoff of the asset just

if some speculators trade on the processed signal. Otherwise, if $\alpha_2^e = 0$, long run price informativeness is equal to short run price informativeness because there is no information production after date 1 and therefore $p_2^* = p_1^*$ with certainty.

[26]This follows from Proposition 6 and the continuity of $\mathcal{E}_2(C_r, C_p)$ in C_r for C_r sufficiently close to zero. The condition $\theta < 1$ is required because for $\theta = 1$, the condition on C_p in Proposition 6 can never be satisfied.

after the arrival of this news. The raw signal s could be, for instance, a signal distributed by information sellers (such as Thomson-Reuters or Ravenpack) that use news analytics to extract information from the news while (s, u) could be a signal produced by buy-side securities analysts after carefully analyzing the implications of the news for a firm. In this context, one could test the implications of our model for price informativeness (and other implications developed in the next section) by considering the effect of a decline in the cost of producing the raw signal (e.g., due to lower access costs to raw data) on the informativeness of stock prices at various dates after news arrival about, say, future earnings (a proxy for V). The model predicts that a decline in the cost of producing the raw signal should make prices more informative shortly after news arrival (say, one day; $t = 1$ in the model) and prices at dates further away from the news (say, one week; $t = 2$ in the model) less informative.

One problem with this approach is that one must take a stand on what are short and long run prices. One way to circumvent this empirical issue is to measure the effect of a reduction in the cost of producing the raw signal on the *average* price observed over some period of time, after the arrival of news. For instance, consider the average price over periods 1 and 2 in our model: $\bar{p}^* = \frac{p_1^* + p_2^*}{2}$. The informativeness of the average price is measured by:

$$\mathcal{E}^{average}(C_r, C_p) = \frac{1}{4} - E[(\tilde{V} - \bar{p}^*)^2] \tag{17}$$

Using the fact that $(V - p_2^*)$ is orthogonal to $(p_2^* - p_1^*)$, we obtain:[27]

$$\mathcal{E}^{average}(C_r, C_p) = 0.75 \times \mathcal{E}_2(C_r, C_p) + 0.25 \times \mathcal{E}_1(C_r, C_p). \tag{18}$$

Thus, as one could expect, the informativeness of the average price over periods 1 and 2 is a weighted average of long run price informativeness and short run price informativeness.

A decline in the cost of producing the raw signal improves short run price informativeness but it can reduce long run price informativeness. Hence, eq.(18) implies that this decline should have a non monotonic effect on the informativeness of the average price over a given period of time. Specifically, numerical simulations show that as C_r declines, the informativeness of the average price first increases (the positive effect on short run price informativeness dominates the negative effect on long run price informativeness)

[27]Indeed, we have $\mathsf{E}[(\tilde{V} - \bar{p}^*)^2] = \mathsf{E}[(\tilde{V} - p_2^*)^2] + 0.25\,\mathsf{E}[(p_2^* - p_1^*)^2]$ because $(V - p_2^*)$ is orthogonal to $p_2^* - p_1^*$. Moreover, for the same reason, we also have $\mathsf{E}[(\tilde{V} - p_1^*)^2] = E[(V - p_2^*)^2] + \mathsf{E}[(p_2^* - p_1^*)^2]$. Hence, $\mathsf{E}[(p_2^* - p_1^*)^2] = \mathsf{E}[(\tilde{V} - p_1^*)^2] - \mathsf{E}[(V - p_2^*)^2]$. We deduce that $\mathsf{E}[(\tilde{V} - \bar{p}^*)^2] = 0.75\,\mathsf{E}[(\tilde{V} - p_2^*)^2] + 0.25\,\mathsf{E}[(V - p_1^*)^2]$, which yields eq.(18).

and then decreases (the negative effect on long run price informativeness dominates). Figure 9 illustrates this pattern for specific parameter values (the informativeness of the average price is given by the green curve). It also shows that the informativeness of the average price when $C_r = 0$ is strictly smaller than the informativeness of the average price when C_r is so large that no speculator buys the raw signal, as implied by Proposition 6.

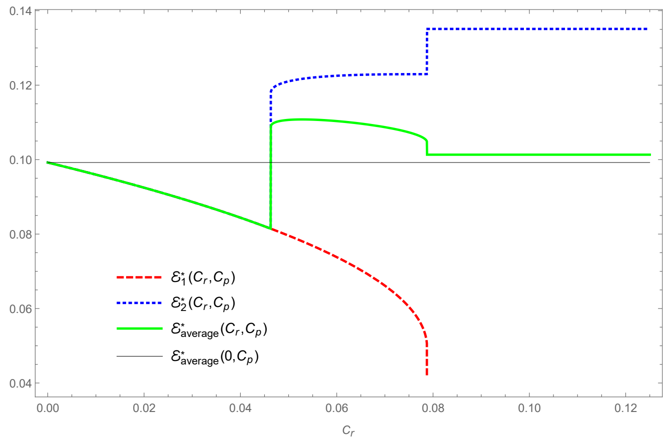


Figure 9: Price informativeness at date $t = 1$ (red dotted line), at date $t = 2$ (blue thick line), and Average price informativeness (green line) as a function of the cost of producing the raw signal, C_r (X-axis). The black line gives the level of Average price informativeness when the cost of producing the raw signal is nil. Other parameters are $\theta = 0.63$ and $C_p = 0.066$

6.2 Price and Trade Patterns

In this section, we analyze in more detail the return and trade patterns induced by speculators' equilibrium behavior. Our goal is to derive additional predictions of our model for the effects of a decrease in the cost of producing the raw signal on the relationships between (i) the trades of speculators at different dates, (ii) past returns and the trades of speculators trading on the processed signal, (iii) future returns and the trades of speculators trading on the raw signal. These predictions could be tested with data on trades by each type of speculators. For instance, discretionary long-short equity hedge funds rely on fundamental analysis of stocks while other hedge funds (or trading desks within these funds) specialize in trading on very high frequency signals (see Pedersen (2015), Chapters 7 and 9). The former trade on processed signals while the latter trade on raw signals according to our terminology.

Corollary 4. *In equilibrium, the covariance between the trades of speculators who buy* *the raw signal (x_1) and the trades of speculators who buy the processed signal (x_2) is:*

$$Cov(x_1, x_2) = \begin{cases} \theta - (1-\theta)\alpha_1^e(\theta, C_r) & \text{if } C_r < \frac{\theta}{8} \quad \text{and} \quad C_p < C_{max}(\theta, \alpha_1^e(\theta, C_r)), \\ 0 & \text{if } C_r > \frac{\theta}{8} \quad \text{or} \quad C_p > C_{max}(\theta, \alpha_1^e(\theta, C_r)), \end{cases}$$

This covariance declines when the cost of producing the raw signal declines and becomes negative if $\theta < \frac{1}{2}$ and $C_r < \frac{\theta^2(2\theta-1)}{2(1-\theta)}$.

Figure 10, Panel A, illustrates Corollary 4. It plots the covariance between trades of speculators at dates 1 and 2 against the reliability of the raw signal, θ, for various values of the cost of producing this signal. This covariance is zero when this cost is so large relative to the reliability of information, θ, that no speculator buys the raw signal ($C_r > \frac{\theta}{8}$) or so small that no speculator buys the processed signal ($C_p > C_{max}(\theta, \alpha_1^e(\theta, C_r))$, which happens for θ large enough, holding C_r constant. For intermediate values of θ, the covariance increases with θ and can be positive or negative. Moreover, holding θ fixed, it decreases as the cost of producing the raw signal declines.

The intuition for Corollary 4 is as follows. The processed and raw signals command trades in the same direction if the raw signal is valid, i.e., when $u = 1$. Instead, when the raw signal is noise, speculators trade on the raw and the processed signals in opposite directions when the price at date 1 reflects the raw signal. Indeed, in this case, speculators who trade on the processed signal correct the noise injected in prices by those trading on the raw signal. Holding θ constant, the probability of the latter event is small when the mass of speculators trading on the raw signal is small (see Figure 2), i.e., the cost of producing the raw signal is large. Hence, for sufficiently high values of C_r, speculators who trade on the raw and the processed signals often trade in the same direction and therefore $Cov(x_1, x_2) > 0$. As the cost of producing the raw signal declines, the likelihood that speculators who trade on the raw signal move prices is higher because more speculators trade on this signal. This effect raises the likelihood that speculators who receive the processed signal trade in a direction opposite to that of speculators who trade on the raw signal and therefore the covariance between the orders of speculators trading at dates 1 and 2 becomes weaker. It can even become negative if the raw signal is sufficiently unreliable (i.e., if $\theta < 1/2$).

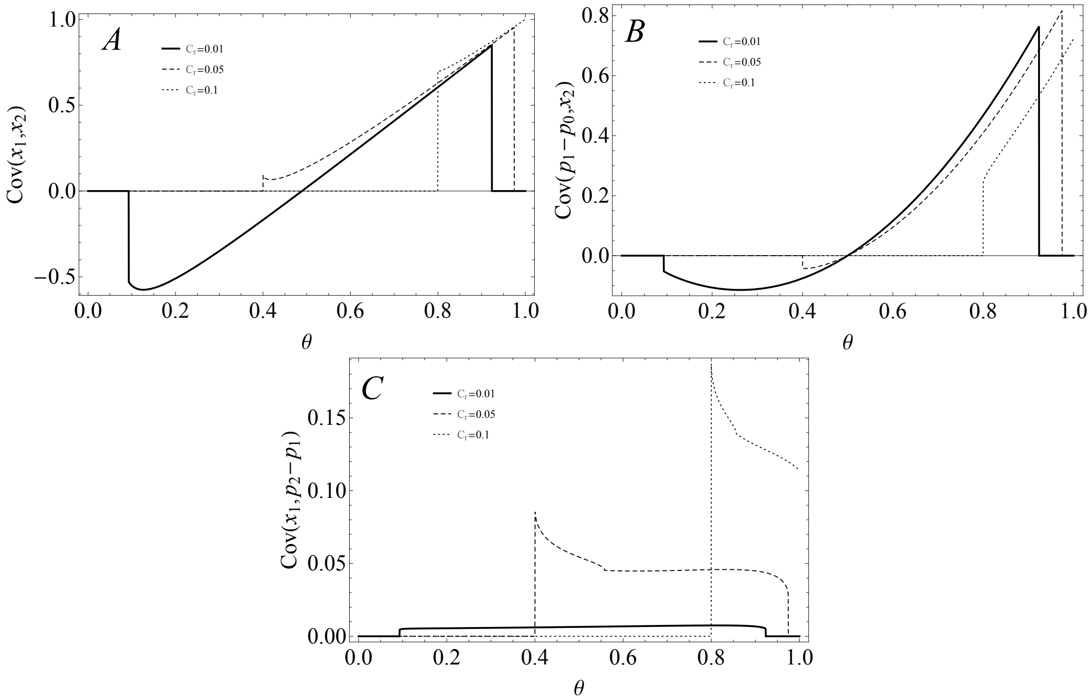


Figure 10: Panel A shows the covariance between speculators' orders at dates 1 and 2 ($Cov(x_1, x_2)$) as a function of θ. Panel B shows the covariance between the return in period 1 ($r_1 = p_1^* - p_0$) and speculators' orders in period 2 as a function of θ. Panel C shows the covariance between speculators' orders in period 1 and the return in period 2 ($p_2^* - p_1^*$) as a function of θ. In each case, various values of C_r are considered: $C_r = 0.1$ (dotted lines), $C_r = 0.05$ (dashed lines), $C_r = 0.01$ (thick lines). In all cases $C_p = 0.02$.

Corollary 5. *In equilibrium, the covariance between the first period return ($r_1 = p_1^* - p_0$) and the trade of speculators who receive the processed signal (x_2) is:*

$$Cov(r_1, x_2) = \begin{cases} \theta(2\theta - 1)\alpha_1^e & \text{if } C_r < \frac{\theta}{8} \quad \text{and} \quad C_p < C_{max}(\theta, \alpha_1^e(\theta, C_r)), \\ 0 & \text{if } C_r > \frac{\theta}{8} \quad \text{or} \quad C_p > C_{max}(\theta, \alpha_1^e(\theta, C_r)), \end{cases}$$

Hence, the trades of speculators who receive the processed signal are negatively correlated with the first period return if and only if $\theta < \frac{1}{2}$. Furthermore, a decline in the cost of producing the raw signal, C_r, raises the absolute value of the covariance between their trade and the first period return.

Figure 10 (Panel B) illustrates this result. Conditional on a price change at date 1, the likelihood that speculators trade against this change after receiving the processed signal increases with the likelihood, $(1 - \theta)$, that the raw signal is noise. This explains why, for $\theta < \frac{1}{2}$, $Cov(r_1, x_2) < 0$. Thus, speculators who trade on the processed signal behave like *momentum traders* when $\theta > \frac{1}{2}$ (the direction of their trades is positively related to the lagged return) and *contrarian traders* (the direction of their trades is negatively related to lagged return) when $\theta < \frac{1}{2}$. Moreover, holding θ constant, the relationship between past returns and their trades becomes stronger when the cost of producing the raw signal declines. Indeed, this decline triggers an increase in the demand for the raw signal and therefore the likelihood that the price at date 1 will adjust to reflect the raw signal.

Corollary 6. *In equilibrium, the covariance between the trade of speculators who receive the raw signal (x_1) and the second period return, $r_2 = p_2^* - p_1^*$, is positive and equal to:*

$$Cov(x_1, r_2) = \begin{cases} 0 \quad \text{when } C_p > C_{max}(\theta, \alpha_1^e), \\ \frac{\theta(1-\alpha_1^e)\alpha_2^e}{2(2-\theta)}, \ \text{when } C_{min}(\theta, \alpha_1^e) \leq C_p \leq C_{max}(\theta, \alpha_1^e), \\ \frac{\theta(1-\alpha_1^e)(1-(1-\theta)(1-\alpha_2^e))}{2(2-\theta)}, \ \text{when } C_p \leq C_{min}(\theta, \alpha_1^e). \end{cases} \tag{19}$$

This covariance decreases when the cost of producing the raw signal declines if (i) $\theta > \frac{\sqrt{2}-1}{\sqrt{2}}$ or (ii) $\theta \leq \frac{\sqrt{2}-1}{\sqrt{2}}$ and $C_r \geq \bar{C}_r(\theta)$, or (iii) $\theta \leq \frac{\sqrt{2}-1}{\sqrt{2}}$ and $C_p \leq \bar{C}_p(\theta)$. In contrast, it always increases when the cost of producing the processed signal declines.

When the cost of producing the raw signal declines, the demand for this signal increases and it becomes increasingly likely that the price at the end of the first period, p_1, reveals the raw signal, s. In this case, speculators receiving the processed signal in the second period can only trade on the component of their signal, (s, u), that is orthogonal to the raw signal (i.e., the innovation in their expectation of the asset payoff due to the observation of u). This effect lowers the covariance between the trade of speculators in the first period and the second period return (see Figure 10, Panel C) because the latter is increasingly determined by factors independent from the raw signal (the realization of u and liquidity traders' orders in the second period).

In contrast, a decline in the cost of producing the processed signal strengthens the covariance between speculators' trade in the first period and the second period return. The reason is that this decrease raises the mass of speculators trading on the processed signal and thereby the likelihood that the return in the second period reveals their signal. As this signal is correlated with the raw signal, s, the predictive power of the trade of speculators using the raw signal for future returns increases.

7 Price Contingent Information Acquisition

In our baseline model, speculators must decide to acquire the processed signal at date 0, i.e., before observing the price at date 1. As the processed signal is delivered only at date 2, another possibility is that speculators wait until observing the realization of the price at date 1 to decide whether or not to buy the processed signal.[28] We have analyzed whether our results were robust to this change in the timing of decisions in our model. For brevity, we report the detailed analysis of this case in Section 3 of the on-line companion appendix.

In this case, the equilibrium demand and the fee for the processed signal vary with the price of the asset at date 1, p_1, because the expected profit from trading on the processed signal is different when the asset price at date 1 reflects the raw signal and when it does not (Corollary 1). However, the implications of the baseline model regarding the effect of a decrease in the cost of producing the raw signal on (i) asset price informativeness and (ii) the relationships between returns and order flows still hold when the decision to buy the processed signal is contingent on the price realized at date 1. In fact, in this case, the negative effect of a reduction in the cost of producing the raw signal on asset price informativeness holds for a broader set of parameters than in our baseline model.

8 Conclusion

Information processing filters out the noise in data but it takes time. Hence, when new data about an asset become available, early signals extracted from these data have a lower precision than later signals. In this paper, we analyze theoretically some implications of this feature of the information production process in financial markets. In particular, we show that a decrease in the cost of producing low precision signals (due, for instance, to lower costs of accessing vast amount of on-line data) can reduce the value of trading on high precision signals because these are available after low precision signals. When this happens, a reduction in the cost of producing low precision signals reduces long run price informativeness, even though it makes prices more informative in the short run (i.e., close to the date at which new data are released).

Our model also predicts that a decline in the cost of producing low precision signals should affect correlations between (i) the trades of speculators trading on low precision signals and those trading on high precision signals, (ii) the trades of speculators trading on high precision signals and past returns, and (iii) the trades of speculators trading on low precision signals and future returns.

[28] This possibility does not seem to have been considered in the literature, even in models featuring multiple trading periods. To our knowledge, researchers have focused on the case in which information acquisition decisions are made ex-ante, i.e., before the realizations of prices, as we do in the baseline version of our model.

Future research could test the implications of our model by considering technological changes that reduce the cost of access to raw data. We believe that recent improvements in technologies to disseminate information in digital form offer many opportunities in this respect.

Our analysis is silent on the welfare effects of a drop in the cost of producing low precision signals. In our model, trading is a zero sum game and therefore information has no social value. In this setting, the total fixed cost of producing signals is a deadweight loss. Thus, a reduction in the cost of producing low precision signals is welfare improving since it reduces this deadweight cost, both directly and indirectly by possibly crowding out investment in the production of more precise signals.[29] However, a complete welfare analysis should also account for possible social gains of more informative prices. In particular, there is growing evidence (see Bond, Edmans, and Goldstein (2012) for a survey) that firms use information in asset prices for their investment decisions. In this case, less informative prices lead to less efficient investment decisions (see, for instance, Dessaint et al.(2016)). Hence, the social benefit of a reduction in information production costs (due to lower access costs to data) should be balanced with the social cost of less efficient decisions for firms due to less informative asset prices. A detailed welfare analysis of this trade off is an interesting venue for future research.

[29] When some investors buy the raw and the processed signals, the total cost of information production is $(C_r + C_p)$ in our model. As the cost of producing the raw signal, C_r, declines, the total cost of information production goes down. Moreover, if the decline in C_r is sufficiently large, the processed signal stops being produced (see Proposition 4) so that the total cost of information production is only C_r.

Appendix A

Proof of Proposition 1.

Step 1: Stock price at date 1. The equilibrium price at date 1 satisfies (see eq.(3)):

$$p_1^*(f_1) = Pr[V = 1|\tilde{f}_1 = f_1] = \frac{Pr[\tilde{f}_1 = f_1|V = 1]Pr[V = 1]}{Pr[\tilde{f}_1 = f_1]}. \tag{20}$$

Speculators buy the asset at date 1 when they observe $s = 1$. Hence, conditional on $V = 1$, aggregate speculators' demand at date 1 is α_1 with probability $(1+\theta)/2$ and $-\alpha_1$ with probability $(1-\theta)/2$. Thus:

$$Pr[\tilde{f}_1 = f_1|V = 1] = (\frac{1+\theta}{2})\phi(f_1 - \alpha_1) + \frac{1-\theta}{2}\phi(f_1 + \alpha_1). \tag{21}$$

Furthermore, by symmetry:

$$Pr[\tilde{f}_1 = f_1] = \frac{1}{2}\phi(f_1 - \alpha_1) + \frac{1}{2}\phi(f_1 + \alpha_1). \tag{22}$$

Substituting (21) and (22) in (20) and using the fact that $Pr[V = 1] = 1/2$, we obtain eq.(4).

Step 2: Speculators' trading strategy at date 1. For a given trade x_1, a speculator's expected profit when he observes signal s is:

$$\pi_1(\alpha, s) = x_1(\mu(s) - \mathsf{E}[p_1|s]).$$

As $p_1^*(f_1) = \mathsf{E}[V|\tilde{f}_1]$ and the market-maker's information set at date 1 is coarser than speculators' information set, we have:

$$\mu(0) \leq p_1^* \leq \mu(1),$$

with a strict inequality when $f_1 \in [-1+\alpha_1, 1-\alpha_1]$ because in this case the order flow at date 1 contains no information (all realizations of the order flow in this interval are equally likely conditional on $V = 0$ or $V = 1$; see Panel A of Figure 2). Therefore:

$$\mu(0) < \mathsf{E}[p_1^*|s] < \mu(1),$$

when $\alpha_1 < 1$. Thus, in this case, it is a strictly dominant strategy for a speculator to buy the asset when $s = 1$ and sell it when $s = 0$. It follows that the equilibrium at date 1 is unique when $\alpha_1 < 1$. When $\alpha_1 \geq 1$, the order flow is fully revealing (see Panel A of Figure 2) and $p_1^*(f_1) = \mu(s)$ for all values of f_1. Hence, a speculator obtains a zero expected profit for all x_1 whether $s = 1$ or $s = 0$. Buying the asset when $s = 1$ and selling the asset when $s = 0$ is then weakly dominant.

Step 3: The expected profit of trading on the raw signal. Suppose that $s = 1$, so that speculators' valuation for the asset after receiving the raw signal is $\mu(1)$. Given their equilibrium strategy, speculators' aggregate demand at date 1 is then α_1. Thus, the aggregate demand for the asset at date 1 is above the threshold $-1 + \alpha_1$. Accordingly, the price at date 1 is either 1/2 if $f_1 \in [-1 + \alpha_1, 1 - \alpha_1]$ or $\mu(1)$ if $f_1 \geq 1 - \alpha_1$ (see Panel B of Figure 2). In the former case, speculators earn a zero expected profit on the raw signal while in the later case, their expected profit is $\mu(1) - 1/2 = \theta/2$. Now we have:

$$Pr(f_1 \in [-1 + \alpha_1, 1 - \alpha_1] \mid s = 1) = Pr(l_1 \in [-1, 1 - 2\alpha_1]) = \max\{1 - \alpha_1, 0\}.$$

Thus, conditional on $s = 1$, speculators' expected profit is $\frac{\theta}{2} \max\{1 - \alpha_1, 0\}$. By symmetry, this is also the case when $s = -1$. Thus, $\bar{\pi}_1(\alpha_1) = \frac{\theta}{2} \max\{1 - \alpha_1, 0\}$.

Proof of Proposition 2.

Step 1. Asset price at date 2. We first derive the equilibrium asset price when speculators behave as described in part 1 of Proposition 2.

Case 1. Suppose first that $p_1 = \mu(1)$. In this case, the market maker knows that $s = 1$. Hence, the remaining uncertainty is about u. If $u = 1$, speculators who receive the processed signal buy the asset at date 2 and, therefore, their total demand for the asset belongs to $[-1 + \alpha_2, f_2^{max}]$. If $u = 0$, these speculators sell the asset since $p_1 > 1/2$ and therefore their total demand for the asset belongs to $[f_2^{min}, 1 - \alpha_2]$. For $\alpha_2 \leq 1$, we have $1 - \alpha_2 > -1 + \alpha_2$. Thus, if $f_2 \in [f_2^{min}, -1 + \alpha_2]$, market makers infer that $u = 0$ and set $p_2^* = \mathsf{E}(V \mid s = 1, u = 0) = 1/2$. Symmetrically if $f_2 \in [1 - \alpha_2, f_2^{max}]$, they infer that $u = 1$ and they set $p_2^* = \mathsf{E}(V \mid s = 1, u = 1) = 1$. Intermediate realizations of f_2 (those in $[-1 + \alpha_2, 1 - \alpha_2]$) are equally likely when $u = 1$ or when $u = 0$. Thus, they convey no information on u. Hence, for these realizations: $p_2^* = \mathsf{E}(V \mid s = 1) = \mu(1)$. For

$\alpha_2 > 1$, the reasoning is unchanged but the intermediate case never occurs. This yields Part 2 of the proposition.

Case 2. When $p_1 = \mu(0)$, the reasoning is symmetric to that followed when $p_1 = \mu(1)$ (Case 1). Part 3 of the proposition follows.

Case 3. Now consider the case in which $p_1 = 1/2$. In this case, the market outcome at date 1 conveys no information to the market maker. Thus, from his viewpoint, there are three possible states at date 2: $(s,u) = (1,1)$, $u = 0$, and $(s,u) = (0,1)$. Given speculators' trading strategy at date 2, the corresponding total demand for the asset at date 2 has the following support: $[-1+\alpha_2, f_2^{max}]$ if $(u,s) = (1,1)$, $[-1,1]$ if $u = 0$, and $[f_2^{min}, 1-\alpha_2]$ if $(s,u) = (0,1)$.

Thus, if $f_2 > 1$, the market maker infers that $(s,u) = (1,1)$ and if $f_2 < -1$, he infers that $(s,u) = (0,1)$. Hence, in the first case $p_2^* = 1$ and in the second case $p_2^* = 0$. Now, consider intermediate realizations for f_2, i.e., $f_2 \in [-1,1]$. First, suppose $f_2 \in [-1, \min\{-1+\alpha_2, 1-\alpha_2\}]$. Such a realization is possible only if $u = 0$ or if $(s,u) = (0,1)$. Thus, in this case:

$$p_2^* = Pr[u = 0 | f_2 \in [-1, \min\{-1+\alpha_2, 1-\alpha_2\}] \times \frac{1}{2}.$$

Now,

$$Pr[u = 0 | f_2 \in [-1, \min\{-1+\alpha_2, 1-\alpha_2\}] = \frac{Pr[f_2 \in [-1, \min\{-1+\alpha_2, 1-\alpha_2\}] | u = 0](1-\theta)}{Pr[f_2 \in [-1, \min\{-1+\alpha_2, 1-\alpha_2\}]},$$

that is

$$Pr[u = 0 | f_2 \in [-1, \min\{-1+\alpha_2, 1-\alpha_2\}] = \frac{2(1-\theta)}{2-\theta}.$$

Thus, for $f_2 \in [-1, \min\{-1+\alpha_2, 1-\alpha_2\}]$, $p_2^* = \frac{1-\theta}{2-\theta}$. The case, in which $f_2 \in [\max\{-1+\alpha_2, 1-\alpha_2\}, 1]$ is symmetric: such a realization of the order flow is possible only if $u = 0$ or if $(s,u) = (1,1)$. Thus, in this case,

$$p_2^* = Pr[(s,u) = (1,1) | f_2 \in [\max\{-1+\alpha_2, 1-\alpha_2\}, 1]] + Pr[u = 0 | f_2 \in [\max\{-1+\alpha_2, 1-\alpha_2\}, 1]]\frac{1}{2}. \tag{23}$$

Using the fact that speculators who receive the processed signal buy the asset if $(s,u) = (1,1)$ and stay put if $u = 0$ (since we are in the case in which $p_1^* = 1/2$), we deduce from eq.(23):

$$p_2^* = \frac{1}{2-\theta}.$$

Finally, realizations of $f_2 \in [\min\{-1+\alpha_2, 1-\alpha_2\}, \max\{-1+\alpha_2, 1-\alpha_2\}]$ are equally likely in each possible state when $p_1 = 1/2$. Thus, observations of f_2 in this range are uninformative and the equilibrium price in this case is $p_2^* = 1/2$. This achieves the proof of Part 4 of the proposition.

Step 2. Speculators' trading strategy at date 2. Let $\mu(s,u)$ be the expected payoff of the asset conditional on the processed signal (s,u). This is the valuation of the asset for the speculators who receive the processed signal at date 2.

Case 1. Suppose that $p_1^* = \mu(1)$. In this case $s = 1$ and either $\mu(1,1) = 1$ or $\mu(1,0) = 1/2$. Moreover, in this case, the equilibrium price of the asset at date 2 is such that:

$$\mu(1,0) \le p_2^* \le \mu(1,1),$$

with a strict inequality when $f_2 \in [-1+\alpha_2, 1]$. This interval is never empty for $\alpha_2 \le 2$. Thus, we can proceed exactly as in the proof of Proposition 1 to show that it is a dominant strategy for speculators receiving the processed signal to (i) buy the asset if their expectation of the value of the asset is $\mu(1,1)$ and $p_1 = \mu(1)$ and (ii) sell the asset if their expectation of the value of the asset is $\mu(1,0)$ and $p_1 = \mu(1)$.

Case 2. Now suppose that $p_1^* = \mu(0)$. In this case, a similar reasoning implies that it is a dominant strategy for the speculators receiving the processed signal to (i) sell the asset if their expectation of the value of the asset is $\mu(0,1)$ and (ii) buy the asset if their expectation of the value of the asset is $\mu(1,0)$.

Case 3. Now consider the case in which $p_1^* = 1/2$ and $u = 1$. In this case, we have:

$$\mu(0,1) \le p_2^* \le \mu(1,1),$$

with a strict inequality for some realizations of f_2. Thus, again, we conclude that it is a dominant strategy for speculators receiving the processed signal to (i) sell the asset if their expectation of the value of the asset is $\mu(0,1)$ and (ii) buy the asset if their expectation of the value of the asset is $\mu(1,1)$.

Case 4. The remaining case is the case in which $p_1^* = 1/2$ and $u = 0$. In this case, a speculator who receive the processed signal expects other speculators to stay put in equilibrium. Suppose that one speculator deviates from this strategy by trading x_2 shares in $[-1,1]$. His effect on aggregate demand is infinitesimal. Hence, he expects $f_2 = l_2$ and therefore he expect f_2 to be uniformly distributed on $[-1, 1]$. Therefore, using the expression for p_2^* when $p_1^* = 1/2$, the speculator expects to trade at:

$$E(p_2^* \mid p_1 = 1/2, f_2 \in [-1,1]) = 1/2 - \frac{\theta(\min\{-1+\alpha_2, 1-\alpha_2\} + \max\{-1+\alpha_2, 1-\alpha_2\})}{4(2-\theta)} = 1/2.$$

As the speculator expects the asset payoff to be $\mu(0,0) = 1/2$, his expected profit is therefore $x_2(\mu(0,0) - E(p_2^* \mid p_1 = 1/2, f_2 \in [-1,1]) = 0$. Thus, the deviation yields a zero expected profit and therefore not trading is weakly dominant for the speculator when $p_1^* = 1/2$ and $u = 0$.

In sum we have shown that the trading strategy described in Part 1 of Proposition 2 is optimal for a speculator who receives the processed signal, if he expects other traders to follow this strategy and if prices at date 2 are given as in Parts 2, 3, and 4 of Proposition 2.

Step 3. Expected profit from trading on the processed signal.

Case 1: $p_1 = \mu(1)$. In this case, a speculator receiving the processed signal buys the asset if $u = 1$ and sells it if $u = 0$. Thus, he makes a profit if and only if $p_2^* = p_1^* = \mu(1)$, i.e., if $f_2 \in [-1+\alpha_2, 1-\alpha-2]$. The likelihood of this event is $\max\{1-\alpha_2, 0\}$ whether $u = 1$ or $u = 0$. Thus, the expected profit of a deep information speculator if $p_1^* = \mu(1)$ is:

$$\pi_2^c(\alpha_2) = \max\{1-\alpha_2, 0\}(\theta \times (1-\mu(1)) + (1-\theta) \times (\mu(1) - 1/2)) = \max\{1-\alpha_2, 0\}\theta(1-\theta).$$

Case 2: $p_1^* = \mu(0)$. The case is symmetric to Case 1 and a speculator receiving the processed signal also earns an expected profit equal to $\pi_2^c(\alpha_2)$.

Case 3: $p_1^* = 1/2$. In this case a speculator trades the asset only if $u = 1$. Suppose first that $s = 1$. Using Parts 2, 3, and 4 of Proposition 2, Table 1 gives the probability of each possible realization for the equilibrium price at date 2 conditional on $\{s, u, p_1^*\} = \{1, 1, 1/2\}$ and the associated profit for a speculator who receives the processed signal (taking into account that speculators buy the asset at date 2 if $(s, u) = (1, 1)$).

We deduce that if $\{s, u, p_1\} = \{1, 1, 1/2\}$, the expected profit of as speculator who receives

Table 1

Equilibrium price at date 2: p_2^*	Prob if $\alpha_2 \leq 1$	Prob if $1 < \alpha_2 \leq 2$	Speculator's profit
0	0	0	1
$\frac{1-\theta}{2-\theta}$	0	0	$\frac{1}{2-\theta}$
$\frac{1}{2}$	$1-\alpha_2$	0	$\frac{1}{2}$
$\frac{1}{2-\theta}$	$\frac{\alpha_2}{2}$	$\frac{(2-\alpha_2)}{2}$	$\frac{1-\theta}{2-\theta}$
1	$\frac{\alpha_2}{2}$	$\frac{\alpha_2}{2}$	0

the processed signal is:

$$\pi_2^{nc}(\alpha_2) = \begin{cases} \frac{\theta}{2(2-\theta)}\left(2-\theta-\alpha_2\right) \text{ if } \alpha_2 \leq 1 \\ \frac{\theta}{2}\frac{1-\theta}{2-\theta}\left(2-\alpha_2\right) \text{ if } 1 < \alpha_2 \leq 2, \\ 0 \text{ if } \alpha_2 > 2, \end{cases} \tag{24}$$

The case in which $(s, u) = (0, 1)$, and $p_1^* = 1/2$ is symmetric and therefore yields the same expected profit for a deep information speculator. Thus, when $p_1^* = 1/2$, the expected profit for a speculator who receives the processed signal is given by $\pi_2^{nc}(\alpha_2)$.

Cases 1 and 2 happen with probability $\alpha_1/2$ each while Case 3 happens with probability $(1-\alpha_1)$. We deduce that the expected profit of a speculator receiving the processed signal is as given by eq.(6).

Proof of Corollary 1 Using the expressions for $\pi_2^{nc}(\alpha_2)$ and $\pi_2^{c}(\alpha_2)$ in Proposition 2, it is direct to show that $\pi_2^{nc}(\alpha_2) < \pi_2^{c}(\alpha_2)$ iff $\alpha_2 < \hat{\alpha}_2(\theta)$ and $\theta \leq 1/2$.

Proof of Corollary 2. It follows directly from Corollary 1 and eq.(8).

Proof of Lemma 1. As explained in the text, $\alpha_2^e = 0$ when $C_p \geq C_{max}$ and $\alpha_2^e \in (\alpha_2^{max}, 2)$ when $0 < C_p < C_{max}$. Let $C_{min}(\theta, \alpha_1)$ be the value of C_p such that $\alpha_2^e = 1$. Thus, C_{min} solves $\pi_2^{gross,a}(\alpha_1, 1) = C_{min}$. Using eq.(6) and the definition of $\pi_2^{gross,a}(\alpha_1, \alpha_2)$, we deduce that

$C_{min}(\theta, \alpha_1) = \frac{\theta}{2}\frac{\theta(1-\theta)(1-\alpha_1)}{2(2-\theta)}$. As $\pi_2^{gross,a}(\alpha_1, \alpha_2)$ decreases continuously in α_2 for $\alpha_2 \in (\alpha_2^{max}, 2)$, we deduce from eq.(12) that $\alpha_2^e \leq 1$ for $C_p > C_{min}$ (case 1) and $\alpha_2^e \geq 1$ for $C_p \leq C_{min}$ (case 2).

Case 1: $C_p > C_{min}$ so that $\alpha_2^e \leq 1$. In this case, using eq.(6) and eq.(12), we deduce that α_2^e solves:

$$\alpha_2^e \bar{\pi}_2(\alpha_1, \alpha_2^e) - C_p = \frac{\theta}{2}\alpha_2 \left[1 - (2\theta - 1)\alpha_1 - \left(\frac{1}{2-\theta} + \left(2(1-\theta) - \frac{1}{2-\theta}\right)\alpha_1\right)\alpha_2\right] - C_p = 0. \tag{25}$$

This equation has two roots in α_2 but only one is larger than α_2^{max}, as required in equilibrium. This root is:

$$\alpha_2^e = \alpha_2^{max}(\theta, \alpha_1)\left(1 + \sqrt{1 - \frac{C_p}{C_{max}(\theta, \alpha_1)}}\right).$$

Case 2: $(C_p \leq C_{min})$ so that $\alpha_2^e \geq 1$. In this case, using eq.(6) and eq.(12), we deduce that α_2^e solves:

$$\alpha_2^e \bar{\pi}_2(\alpha_1, \alpha_2^e) - C_p = \frac{\theta(1-\theta)}{2(2-\theta)}(1-\alpha_1)\alpha_2(2-\alpha_2) - C_p = 0. \tag{26}$$

This equation again has two roots in α_2 but only one is larger than 1 (as required). This root is:

$$\alpha_2^e = 1 + \sqrt{1 - \frac{C_p}{C_{min}(\theta, \alpha_1)}}.$$

Proof of Proposition 3. As C_r affects α_2^e only through its effect on α_1^e, we have:

$$\frac{\partial \alpha_2^e}{\partial C_r} = \left(\frac{\partial \alpha_2^e}{\partial \alpha_1}\right)\left(\frac{\partial \alpha_1^e}{\partial C_r}\right). \tag{27}$$

It is immediate from Lemma 2 that $\frac{\partial \alpha_1^e}{\partial C_r} \leq 0$. Thus, eq.(27) implies that $\frac{\partial \alpha_2^e}{\partial C_r} \geq 0$ iff $\frac{\partial \alpha_2^e}{\partial \alpha_1} < 0$. Thus, in the rest of this proof, we sign $\frac{\partial \alpha_2^e}{\partial \alpha_1}$.

Remember that α_2^e solves:

$$\pi_2^{gross,a}(\alpha_1, \alpha_2^e) = \alpha_2^e \bar{\pi}_2(\alpha_1, \alpha_2^e) = C_p.$$

Thus, using the implicit function theorem, we have

$$\frac{\partial \alpha_2^e}{\partial \alpha_1} = -\frac{\frac{\partial}{\partial \alpha_1}\left[\alpha_2 \bar{\pi}_2(\alpha_1, \alpha_2)\right]_{\alpha_2 = \alpha_2^e}}{\frac{\partial}{\partial \alpha_2}\left[\alpha_2 \bar{\pi}_2(\alpha_1, \alpha_2)\right]_{\alpha_2 = \alpha_2^e}}. \tag{28}$$

For $C_p < C_{max}$, $\alpha_2^e > \alpha_2^{max}$. Thus, we have:

$$\frac{\partial}{\partial \alpha_2} \left[\alpha_2 \bar{\pi}_2(\alpha_1, \alpha_2)\right]_{\alpha_2 = \alpha_2^e} < 0. \tag{29}$$

We deduce from eq.(28) that $\frac{\partial \alpha_2^e}{\partial \alpha_1} > 0$ iff $\frac{\partial}{\partial \alpha_1} \left[\alpha_2 \bar{\pi}_2(\alpha_1, \alpha_2)\right]_{\alpha_2 = \alpha_2^e} < 0$.

Case 1: $\theta > 1/2$ or $C_p < C_{min}(\theta, \alpha_1)$. If $\theta > 1/2$, we deduce from Corollary 2 that the expected profit of a speculator who trades on the processed signal, $\bar{\pi}_2$, decreases with α_1. If $C_p < C_{min}(\theta, \alpha_1)$, we deduce from Proposition 1 that $\alpha_2^e > 1$. Therefore, using Corollary 2 again, $\bar{\pi}_2$, decreases with α_1. Hence, for $\theta > 1/2$ or $C_p < C_{min}(\theta, \alpha_1)$, we have:

$$\frac{\partial[\alpha_2 \bar{\pi}_2(\alpha_1, \alpha_2^e)]}{\partial \alpha_1} < 0.$$

We deduce from eq.(28)and eq.(29) that if $\theta > 1/2$ or $C_p < C_{min}(\theta, \alpha_1)$ then $\frac{\partial \alpha_2^e}{\partial \alpha_1} < 0$ and therefore $\frac{\partial \alpha_2^e}{\partial C_r} > 0$.

Case 2: $\theta \leq 1/2$ and $C_{min}(\theta, \alpha_1) < C_p < C_{max}(\theta, \alpha_1)$. Using Corollary 2, we deduce that the expected profit of a speculator who trades on the processed signal, $\bar{\pi}_2$, increases with α_1 iff $\alpha_2^e(\alpha_1) < \hat{\alpha}_2(\theta)$. Thus, if this condition is satisfied then $\frac{\partial}{\partial \alpha_1} \left[\alpha_2 \bar{\pi}_2(\alpha_1, \alpha_2)\right]_{\alpha_2 = \alpha_2^e} > 0$ and therefore $\frac{\partial \alpha_2^e}{\partial \alpha_1} > 0$. Thus, in this case, $\frac{\partial \alpha_2^e}{\partial C_r} < 0$. The rest of the proof consists in showing that the conditions (i) $\theta < \frac{\sqrt{2}-1}{\sqrt{2}}$, (ii) $C_r < \bar{C}_r(\theta)$, and (iii) $C_p > \bar{C}_p(\theta)$ are necessary and sufficient for $\alpha_2^e(\alpha_1) < \hat{\alpha}_2(\theta)$. For brevity, we provide the proof of this result in Section 4 of the on-line appendix. As $\frac{\sqrt{2}-1}{\sqrt{2}} < 1/2$, the proposition follows.

Proof of Proposition 4. When $\theta > \frac{\sqrt{2}-1}{\sqrt{2}}$, we show in Section 6 of the on-line appendix that $C_{max}(\theta, \alpha_1)$ decreases with α_1. Moreover, using eq.(14), we obtain $C_{max}(\theta, 1) = \frac{\theta(1-\theta)}{4}$ and $C_{max}(\theta, 0) = \frac{\theta(2-\theta)}{8}$. Thus, for each $C_p \in [\frac{\theta(1-\theta)}{4}, \frac{\theta(2-\theta)}{8}]$, there exists a unique $\alpha_1^c(\theta, C_p)$ such that:

$$C_p = C_{max}(\theta, \alpha_1^c).$$

Moreover, for $\alpha_1^e > \alpha_1^c$, $C_p < C_{max}(\theta, \alpha_1^e)$ while for $\alpha_1^e < \alpha_1^c$, $C_p > C_{max}(\theta, \alpha_1^e)$. We deduce from Lemma 1, that for $\alpha_1^e > \alpha_1^c$, $\alpha_2^e(\theta, \alpha_1^e) > \alpha_2^{max}$ while for $\alpha_1^e < \alpha_1^c$, $\alpha_2^e(\theta, \alpha_1^e) = 0$. The proposition follows by defining $\widehat{C}_r$ as the value of C_r such that $\alpha_1^e(\theta, \widehat{C}_r) = \alpha_1^c(\theta, C_p)$.

Proof of Corollary 3. Using Proposition 1 (or Figure 2), we obtain that:

$$\mathcal{E}_1(C_r, C_p) = \begin{cases} 0 \quad \text{if } C_r \geq \frac{\theta}{8}, \\ \frac{\alpha_1^e(\theta, C_r)\theta^2}{4} \quad \text{if } C_r \leq \frac{\theta}{8}, \end{cases} \tag{30}$$

and

$$\mathcal{E}_2(C_r, C_p) = \begin{cases} \mathcal{E}_1(C_r, C_p) & \text{if } C_p \geq C_{max}(\theta, \alpha_1^e), \\ \frac{\theta}{4}\left[1 - (1-\alpha_1^e)\left(1 - \frac{\alpha_2^e}{2-\theta}\right) - (1-\theta)\alpha_1^e(1-\alpha_2^e)\right] & \text{if } C_{min}(\theta, \alpha_1^e) \leq C_p \leq C_{max}(\theta, \alpha_1^e), \\ \frac{\theta}{4}\left[1 - \frac{1-\theta}{2-\theta}(1-\alpha_1^e)(2-\alpha_2^e)\right] & \text{if } C_p \leq C_{min}(\theta, \alpha_1^e), \end{cases} \tag{31}$$

where to simplify notations we have omitted the arguments of functions α_2^e and α_1^e. As α_1^e does not depend on the cost of producing the processed signal, we deduce from eq.(30) that price informativeness at date 1 is not affected by a change in C_p.

For $C_p < C_{max}(\theta, \alpha_1^e)$, it is immediate from eq.(31) that price informativeness at date 2 increases with α_2^e. As α_2^e declines when C_p decreases, we deduce that price informativeness at date 2 increases when C_p declines for $C_p < C_{max}(\theta, \alpha_1^e)$. For $C_p > C_{max}(\theta, \alpha_1^e)$, price informativeness at date 2 is equal to price informativeness at date 1 and therefore independent of C_p.

Proof of Proposition 5.

Part 1: Effect of C_r on short run price informativeness. We know from Proposition 2 that α_1^e weakly increases when C_r decreases. Hence, we deduce from eq.(30) that $\mathcal{E}_1(C_r, C_p)$ weakly decreases when C_r decreases.

Part 2: Effect of C_r on long run price informativeness. Suppose that $C_p < C_{min}(\theta, \alpha_1^e)$. In this case, $\alpha_2^e \geq 1$ (Lemma 1). Using eq.(6) and eq.(31), we have:

$$\mathcal{E}_2(C_r, C_p) = \frac{\theta}{4} - \frac{1}{2}\bar{\pi}_2(\alpha_1^e, \alpha_2^e), \tag{32}$$

where we omit the arguments of functions α_1^e and α_2^e to simplify notations. Now, as $\alpha_2^e > 0$, in equilibrium, $\alpha_2^e \bar{\pi}_2^e = C_p$ (see eq.(11)). Thus, we deduce from (32) that:

$$\mathcal{E}_2 = \frac{\theta}{4} - \frac{1}{2}\frac{C_p}{\alpha_2^e}. \tag{33}$$

As $C_p < C_{min}(\theta, \alpha_1^e)$, we deduce from the analysis of Case 1 in the proof of Proposition 3 that α_2^e decreases when C_r decreases. Hence, from eq.(33), we deduce that if $C_p < C_{min}(\theta, \alpha_1^e)$ then $\mathcal{E}_2(C_r, C_p)$ decreases when C_r decreases.

Proof of Proposition 6. Note that $\alpha_1^e = 0$ for all $C_r \geq \frac{\theta}{8}$. Thus, $\mathcal{E}_2(C_r, C_p) = \mathcal{E}_2(\frac{\theta}{8}, C_p)$ for $C_r \geq \frac{\theta}{8}$. We denote the difference in price informativeness at date 2 when $C_r = 0$ and when $C_r \geq \frac{\theta}{8}$, for a given C_p, by $\Delta\mathcal{E}_2(C_p)$. That is:

$$\Delta\mathcal{E}_2(C_p) \equiv \mathcal{E}_2(\frac{\theta}{8}, C_p) - \mathcal{E}_2(0, C_p) \tag{34}$$

Observe that:

$$C_{max}(\theta, 0) = \frac{\theta(2-\theta)}{8} > C_{min}(\theta, 0) = \frac{\theta(1-\theta)}{2(2-\theta)} > C_{max}(\theta, 1) = \frac{\theta(1-\theta)}{4} > C_{min}(\theta, 1) = 0, \tag{35}$$

Case 1. First, consider the case in which $C_p \in [0, C_{max}(\theta, 1)]$. In this case, using Lemma 1, and eq.(35), we obtain that if $C_r = 0$ then

$$\alpha_2^e = \frac{1}{2}\left(1 + \sqrt{1 - \frac{C_p}{C_{max}(\theta, 1)}}\right) < 1,$$

because $\alpha_1^e = 1$ when $C_r = 0$ and in this case $\alpha_2^{max}(\theta, 0) = 12)$. Similarly, if $C_r \geq \frac{\theta}{8}$ then

$$\alpha_2^e = 1 + \sqrt{1 - \frac{C_p}{C_{max}(\theta, 0)}} > 1,$$

because $\alpha_1^e = 0$ if $C_r \geq \frac{\theta}{8}$ and $C_p < C_{max}(\theta, 1) < C_{min}(\theta, 0)$. Hence, using eq.(31), eq.(34), and eq.(35), we obtain

$$\Delta\mathcal{E}_2(C_p) = \frac{\theta}{4}\left[\frac{(1-\theta)}{2}\left(1 - \sqrt{1 - \frac{C_p}{C_{max}(\theta, 1)}}\right) - (\frac{1-\theta}{2-\theta})\left(1 - \sqrt{1 - \frac{C_p}{C_{max}(\theta, 0)}}\right)\right].$$

Using the expressions for $C_{max}(\theta,0)$ and $C_{max}(\theta,1)$, we deduce after some algebra that:

$$\frac{4}{\theta}\frac{\partial\Delta\mathcal{E}_2}{\partial C_p}=\frac{1}{\theta}\left(\frac{1}{\sqrt{1-\frac{C_p}{C_{max}(\theta,1)}}}-\frac{1}{\sqrt{1-\frac{C_p}{C_{max}(\theta,0)}}}\right).$$

This is always positive since $C_{max}(\theta,0)>C_{max}(\theta,1)$. Thus, for $C_p\in[0,C_{max}(\theta,1)]$, $\Delta\mathcal{E}_2(C_p)$ increases with C_p. As $\Delta\mathcal{E}_2(0)=0$ and $\Delta\mathcal{E}_2(C_{max}(\theta,1))>0$, we obtain that $\Delta\mathcal{E}_2(C_p)>0$ when $C_p\in[0,C_{max}(\theta,1)]$.

Case 2. Now, consider the case in which $C_p\in[C_{max}(\theta,1),C_{min}(\theta,0)]$. In this case, using Lemma 1, we obtain that if $C_r\geq\frac{\theta}{8}$ then

$$\alpha_2^e=1+\sqrt{1-\frac{C_p}{C_{min}(\theta,0)}}>1,$$

because α_1^e=0 when $C_r\geq\frac{\theta}{8}$. Similarly, if $C_r=0$ then

$$\alpha_2^e=0,$$

because α_1^e=1 when $C_r=0$ and $C_p\geq C_{max}(\theta,1)$. Hence, using eq.(31) and eq.(34), we obtain

$$\Delta\mathcal{E}_2(C_p)=\frac{\theta}{4}\left[1-\theta-\frac{1-\theta}{2-\theta}\left(1-\sqrt{1-\frac{C_p}{C_{min}(\theta,0)}}\right)\right]=\frac{\theta}{4}\left[\frac{(1-\theta)^3}{(2-\theta)^2}\sqrt{1-\frac{C_p}{C_{min}(\theta,0)}}\right]>0.$$

Case 3. Finally suppose that $C_p\in[C_{min}(\theta,0),C_{max}(\theta,0)]$. In this case, using Lemma 1, we obtain that if $C_r\geq\frac{\theta}{8}$ then

$$\alpha_2^e=\left(1-\frac{\theta}{2}\right)\left(1+\sqrt{1-\frac{C_p}{C_{max}(\theta,0)}}\right)<1,$$

because $\alpha_1^e=0$ if $C_r\geq\frac{\theta}{8}$ and in this case $\alpha_2^{max}(\theta,0)=(1-\frac{\theta}{2})$. Similarly, if $C_r=0$ then

$$\alpha_2^e=0,$$

because α_1^e=1 when $C_r=0$ and $C_p\geq C_{max}(\theta,1)$ (which is the case since $C_{max}(\theta,1)<C_{min}(\theta,0)<C^p$). Hence, using eq.(31) and eq.(34), and the fact that $\alpha_1^e=1$ if $C_r=0$,

we obtain that

$$\Delta\mathcal{E}_2 = \frac{\theta}{4}\left[\frac{1}{2} - \theta + \frac{1}{2}\sqrt{1 - \frac{8}{\theta(2-\theta)}C_p}\right],$$

which is positive if $C_p \leq \hat{C}_p(\theta)$ where $\hat{C}_p(\theta) = (1-(2\theta-1)^2)\frac{\theta(2-\theta)}{8}$.

Proof of Corollary 4. Using the first parts of Propositions 1 and 2, we deduce that:

$$x_1 = \mathbb{I}_{s=1} - \mathbb{I}_{s=0}, \text{ with } s = u \times V + (1-u) \times \epsilon, \tag{36}$$

$$x_2 = u \times [\mathbb{I}_{V=1} - \mathbb{I}_{V=0}] + (1-u) \times [\mathbb{I}_{p1=(1-\theta)/2} - \mathbb{I}_{p1=(1+\theta)/2}], \tag{37}$$

where $\mathbb{I}$ denotes the indicator function, which is equal to one when the statement in brackets holds. As $\mathbb{E}[x_1] = \mathbb{E}[x_2] = 0$, we deduce from eq. (36) and eq.(37) that:

$$\begin{aligned}
Cov(x_1,x_2) = \mathsf{E}[x_1x_2] &= \frac{1}{2}\,\mathsf{E}[x_2|s=1] - \frac{1}{2}\,\mathsf{E}[x_2|s=0],\\
&= \frac{\theta}{2}\,\mathsf{E}\left[x_2|V=1,u=1\right] + \frac{1}{2}(1-\theta)\frac{\alpha_1^e}{Q}\,\mathsf{E}\left[x_2|\epsilon=1,u=0,p_1=\frac{1+\theta}{2}\right],\\
&- \frac{\theta}{2}\,\mathsf{E}\left[x_2|V=0,u=1\right] - \frac{1}{2}(1-\theta)\frac{\alpha_1^e}{Q}\,\mathsf{E}\left[x_2|\epsilon=0,u=0,p_1=\frac{1-\theta}{2}\right],\\
&= \theta - (1-\theta)\alpha_1^e.
\end{aligned}$$

As α_1^e increases when C_r declines, we deduce that $Cov(x_1,x_2)$ decreases when C_r decreases. Moreover, $Cov(x_1,x_2) < 0$ iff:

$$\alpha_1^e(\theta,C_r) \geq \frac{\theta}{1-\theta}.$$

Substituting $\alpha_1^e(\theta,C_r)$ by its expression in eq.(15), we deduce that $Cov(x_1,x_2) < 0$ iff $\theta < 1/2$ and $C_r < \frac{\theta^2(2\theta-1)}{2(1-\theta)}$.

Proof of Corollary 5. Using the second part of Proposition 1 and the first part of Proposition 2, we deduce that:

$$p_1 = \frac{1}{2} + \frac{\theta}{2}\mathbb{I}_{f_1>1-\alpha_1^e} - \frac{\theta}{2}\mathbb{I}_{f_1<-1+\alpha_1^e} \tag{38}$$

$$x_2 = U \times [\mathbb{I}_{V=1} - \mathbb{I}_{V=0}] + (1-u) \times [\mathbb{I}_{p1=(1-\theta)/2} - \mathbb{I}_{p_1^*=(1+\theta)/2}]. \tag{39}$$

As $\mathsf{E}[x_2] = 0$ and $\mathsf{E}[p_1] = 1/2$, we deduce from (38) and (39) that:

$$
\begin{aligned}
Cov(p_1, x_2) = \mathsf{E}[(p_1^* - 1/2)x_2] &= \frac{\theta\alpha_1^e}{4}\left\{\mathsf{E}\left[x_2 | s = 1, p_1^* = \frac{1+\theta}{2}\right] - \mathsf{E}\left[x_2 | s = 0, p_1 = \frac{1-\theta}{2}\right]\right\} \\
&= \frac{\theta^2}{4}\alpha_1^e\,\mathsf{E}\left[x_2 | V = 1, u = 1, p_1^* = \frac{1+\theta}{2}\right] + \frac{\theta(1-\theta)}{4}\alpha_1^e\,\mathsf{E}\left[x_2 | \epsilon = 1, u = 0, p_1^* = \frac{1+\theta}{2}\right] \\
&- \frac{\theta^2}{4}\alpha_1^e\,\mathsf{E}\left[x_2 | V = 0, u = 1, p_1^* = \frac{1-\theta}{2}\right] - \frac{\theta(1-\theta)}{4}\alpha_1^e\,\mathsf{E}\left[x_2 | \epsilon = 0, u = 0, p_1^* = \frac{1-\theta}{2}\right] \\
&= \theta(2\theta - 1)\alpha_1^e.
\end{aligned}
$$

As α_1^e increases when C_r declines, we deduce that $|Cov(p_1, x_2)|$ increases when C_r decreases.

Proof of Corollary 6.

We first compute the expression for $Cov(x_1, r_2)$ given in eq.(19). As $\mathsf{E}[x_1] = 0$,

$$
Cov(x_1, r_2) = \mathsf{E}[(p_2^* - p_1^*)x_1] - \mathsf{E}[p_2^* - p_1^*]\,\mathsf{E}[x_1] = \mathsf{E}[(p_2^* - p_1^*)x_1]. \tag{40}
$$

Now:

$$
\begin{aligned}
\mathsf{E}[p_1^* x_1] &= \frac{1}{2}(\mathsf{E}[p_1^* x_1 | s = 1] + \mathsf{E}[p_1^* x_1 | s = 0]) = \frac{1}{2}(\mathsf{E}[p_1^* | s = 1] - \mathsf{E}[p_1 | s = 0]) \\
&= \frac{1}{2}\left((1 - \alpha_1^e)\frac{1}{2} + \alpha_1^e\frac{1+\theta}{2}\right) - \frac{1}{2}\left((1 - \alpha_1^e)\frac{1}{2} + \alpha_1^e\frac{1-\theta}{2}\right) \\
&= \frac{\theta\alpha_1^e}{2}.
\end{aligned} \tag{41}
$$

Similarly , we have that:

$$
\mathsf{E}[p_2^* x_1] = \frac{1}{2}(\mathsf{E}[p_2^* | s = 1] - \mathsf{E}[p_2^* | s = 0]). \tag{42}
$$

We first compute $E[p_2^* | s = 1]$. We have:

$$
\mathsf{E}[p_2^* | s = 1] = \alpha_1^e\,\mathsf{E}\left[p_2 \middle| s = 1, p_1^* = \frac{1+\theta}{2}\right] + (1 - \alpha_1^e)\,\mathsf{E}\left[p_2^* \middle| s = 1, p_1^* = \frac{1}{2}\right]. \tag{43}
$$

The event $p_1 = \frac{1+\theta}{2}$ implies that $s = 1$. Thus,

$$
\mathsf{E}\left[p_2^* \middle| s = 1, p_1^* = \frac{1+\theta}{2}\right] = \mathsf{E}\left[p_2 \middle| p_1^* = \frac{1+\theta}{2}\right] = p_1^* = \frac{1+\theta}{2},
$$

where the third equality follows from the fact that the equilibrium price is a martingale. Moreover, using Proposition 2, we deduce that if $\alpha_2^e < 1$:

$$\begin{aligned}
\mathsf{E}\left[p_2^* \middle| s=1, p_1^*=\frac{1}{2}\right] &= \frac{1}{2}(1-\theta)\alpha_2^e \times \frac{1-\theta}{2-\theta} + (1-\alpha_2^e) \times \frac{1}{2} + \frac{1}{2}\alpha_2^e \times \frac{1}{2-\theta} + \frac{1}{2}\theta\alpha_2^e \times 1 \\
&= \frac{1}{2} + \alpha_2^e\left(-\frac{1}{2} + \frac{(1-\theta)^2}{2(2-\theta)} + \frac{1}{2(2-\theta)} + \frac{\theta}{2}\right) \\
&= \frac{1}{2} + \frac{\theta\alpha_2^e}{2(2-\theta)}.
\end{aligned}$$

and if $\alpha_2^e \geq 1$,

$$\begin{aligned}
\mathsf{E}\left[p_2^* \middle| s=1, p_1^*=\frac{1}{2}\right] &= (1-\theta)\frac{1}{2} + \theta\left[\frac{\alpha_2^e}{2} \times 1 + \left(1-\frac{\alpha_2^e}{2}\right) \times \frac{1}{2-\theta}\right] \\
&= \frac{1}{2} + \theta\left[\frac{\alpha_2^e}{2}\left(1-\frac{1}{2-\theta}\right) + \frac{1}{2-\theta} - \frac{1}{2}\right] \\
&= \frac{1}{2} + \theta\left[\frac{\theta}{2(2-\theta)} + \frac{1-\theta}{2(2-\theta)}\alpha_2^e\right] \\
&= \frac{1}{2} + \frac{\theta}{2(2-\theta)}[1+(1-\theta)(\alpha_2^e-1)].
\end{aligned}$$

We deduce from these expressions and eq.(43) that:

$$\mathsf{E}[p_2^*|s=1] = \begin{cases} \frac{1+\theta\alpha_1^e}{2} + \frac{\theta(1-\alpha_1^e)\alpha_2^e}{2(2-\theta)} & \text{if } \alpha_2^e \leq 1, \\ \frac{1+\theta\alpha_1^e}{2} + \frac{\theta(1-\alpha_1^e)}{2(2-\theta)}[1-(1-\theta)(1-\alpha_2^e)] & \text{if } \alpha_2^e > 1. \end{cases} \tag{44}$$

Proceeding in a similar way, we obtain after some algebra that

$$\mathsf{E}[p_2^*|s=1] = \begin{cases} \frac{1-\theta\alpha_1^e}{2} - \frac{\theta\alpha_2^e}{2(2-\theta)} & \text{if } \alpha_2^e \leq 1, \\ \frac{1-\theta\alpha_1^e}{2} - \frac{\theta}{2(2-\theta)}[1-(1-\theta)(1-\alpha_2^e)] & \text{if } \alpha_2^e > 1. \end{cases} \tag{45}$$

After some algebra, we deduce from equations (40), (41), (42), (44), and (45) that:

$$Cov(x_1, r_2) = \begin{cases} \frac{\theta(1-\alpha_1^e)\alpha_2^e}{2(2-\theta)} & \text{if } \alpha_2^e \leq 1, \\ \frac{(1-\alpha_1^e)\theta}{2(2-\theta)}[1-(1-\theta)(1-\alpha_2^e)] & \text{if } \alpha_2^e > 1, \end{cases} \tag{46}$$

which is equivalent to eq.(19) because $\alpha_2^e > 1$ iff $C_p < C_{min}(\theta, \alpha_1^e)$ and $\alpha_2^e \leq 1$ iff $C_{min}(\theta, \alpha_1^e) \leq C_p \leq C_{max}(\theta, \alpha_1^e)$.

As α_2^e decreases with C_p and α_1^e does not depend on C_p, it is immediate from eq.(19) that

$Cov(x_1, r_2)$ increases when C_p decreases. Moreover, if (i) $\theta > \frac{\sqrt{2}-1}{\sqrt{2}}$ or (ii) $\theta \leq \frac{\sqrt{2}-1}{\sqrt{2}}$ and $C_r \geq \bar{C}_r(\theta)$, or (iii) $\theta \leq \frac{\sqrt{2}-1}{\sqrt{2}}$ and $C_p \leq \bar{C}_p(\theta)$ then α_2^e decreases when C_r decreases. Thus, as α_1^e increases when C_r decreases, we deduce that $Cov(x_1, r_2)$ decreases when C_r decreases if i) $\theta > \frac{\sqrt{2}-1}{\sqrt{2}}$ or (ii) $\theta \leq \frac{\sqrt{2}-1}{\sqrt{2}}$ and $C_r \geq \bar{C}_r(\theta)$, or (iii) $\theta \leq \frac{\sqrt{2}-1}{\sqrt{2}}$ and $C_p \leq \bar{C}_p(\theta)$.

References

[1] Admati, A. and P. Pfleiderer, (1986). "A Monopolistic Market for Information," *Journal of Economic Theory* 39, 400-438.

[2] Bai, J., Philippon, T. and A. Savov (2015). "Have Financial Markets Become More Informative?" forthcoming, *Journal of Financial Economics.*

[3] Bond, P., Edmans, A., Goldstein, I., (2012). The real effects of financial markets. *Annual Review of Financial Economics* 4, 339–360.

[4] Boulatov, A., T. Hendershott, and D. Livdan (2013). "Informed Trading and Portfolio Returns," *Review of Economic Studies* 80, 35-72.

[5] Brunnermeier M. (2005). "Information leakage and market efficiency", *Review of Financial Studies* 18, 417-457.

[6] Cespa. G. (2008). "Information Sales and Insider Trading with Long Lived Information". *Journal of Finance*, 73, 639-672

[7] Chen H., P. De, Y. Hu, and B. Hwang (2014). "Wisdom of Crowds: The value of stock opinions transmitted through social medias." *Review of Financial Studies*, 27, 1367-1403.

[8] Dessaint, O., Foucault, T., Frésard, L., and , A., Matray, (2012). Ripple effects of noise on investment. Available at: http://dx.doi.org/10.2139/ssrn.2707999.

[9] Engelberg J., Reed, A., and Ringgenberg, M.(2012). "How are shorts informed? Short sellers, news, and information processing," *Journal of Financial Economics*, 260-278.

[10] Froot K., Scharfstein D. and Stein J. (1992). "Herd on the street: informational efficiencies in a market with short-term speculation", *Journal of Finance* 47, 1461-1484.

[11] Grossman, S. and J. Stiglitz (1980). "On the impossibility of informationally efficient markets," *American Economic Review*, 70, 393-408.

[12] Glosten L. and Milgrom P.(1985). "Bid, ask and transaction prices in specialist market with heterogeneously informed trader." *Journal of Financial Economics*, 13, 71-100.

[13] Hirshleifer D., A. Subrahmanyam, and S. Titman (1994). "Security analysis and trading patterns when some investors receive information before others", *Journal of Finance* 49, 1665-1698.

[14] Holden C. and A. Subrahmanyam (2002). "News Events, Information Acquisition, and Stock Price Behavior," *Journal of Business* 75, 1-32.

[15] Holden C. and A. Subrahmanyam (1996). “Risk aversion, liquidity, and endogenous short horizons,” *Review of Financial Studies*, 9, 691-722.

[16] Kyle A. (1985). ”Continuous auctions and insider trading”, *Econometrica* 53, 1315-1336.

[17] Lee S. (2013). ”Active investment, liquidity externalities, and markets for information”, *Working paper,* available at http://dx.doi.org/10.2139/ssrn.1300706.

[18] Pedersen L. (2015), Efficiently inefficient, Princeton University Press.

[19] Peress, J. (2010). ”The trade-off between risk sharing and information production in financial markets.” *Journal of Economic Theory* 145, 124-155.

[20] SEC (2009). “Interactive Data to Improve Financial Reporting”, 1-206.

[21] Shapiro, C. and H. Varian (1999). “Information rules,” Harvard Business School Press.

[22] Veldkamp, L. (2006a). “Media Frenzies in Markets for Financial Information,” *American Economic Review*, 96, 577-601.

[23] Veldkamp, L. (2006b). “Information Markets and the Comovement of Asset Prices,” *Review of Economic Studies*, 73, 823-845.

[24] Veldkamp, L. (2011). “Information choice in macroeconomics and finance,” *Princeton University Press*, 167 pages.

[25] Verrecchia, R. E. (1982). “Information Acquisition in a Noisy Rational Expectations Economy.” *Econometrica*, 50, 1415-30.

[26] Weller, B. (2016). “Efficient prices at any cost: does algorithmic trading deter information acquisition?” available at: http://dx.doi.org/10.2139/ssrn.2662254.