Data Abundance and Asset Price Informativeness

Jérôme Dugast [1] Thierry Foucault[2]

[1]Luxemburg School of Finance

[2]HEC Paris

NUS

"Trading on Big Data"

THE WALL STREET JOURNAL.

Why now ?

- **Improvements in information technologies have reduced the cost of accessing vast amount of "raw" (unstructured) data (e.g., text, images, or audio records)** = Data Abundance
- ⇒ **Increase in supply and demand of trading signals based on raw data in real time (e.g., news reports, press releases, tweets, Facebook pages, satellite images, voice analysis etc.)**

Supply and Demand of Big Data for Trading

- **Supply :** *Everyday iSentium [...] analyzes one million tweets from traders, investors, and market commentators to try to find out whether the sentiment for a particular stock is high or low. The answer is simple : either a +1 or a -1 [...]. Yet, a handful of banks hedge funds and high frequency traders have signed up [...] at a cost of* $15,000 *per month per stock.*"in "Firms analyze tweets to gauge market sentiment," WSJ, July 6, 2015.

Supply and Demand of Big Data for Trading

- **Supply :** *Everyday iSentium [...] analyzes one million tweets from traders, investors, and market commentators to try to find out whether the sentiment for a particular stock is high or low. The answer is simple : either a +1 or a -1 [...]. Yet, a handful of banks hedge funds and high frequency traders have signed up [...] at a cost of* $15,000 *per month per stock.*"in "Firms analyze tweets to gauge market sentiment," WSJ, July 6, 2015.
- **Demand :** *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.

Supply and Demand of Big Data for Trading

- **Supply :** *Everyday iSentium [...] analyzes one million tweets from traders, investors, and market commentators to try to find out whether the sentiment for a particular stock is high or low. The answer is simple : either a +1 or a -1 [...]. Yet, a handful of banks hedge funds and high frequency traders have signed up [...] at a cost of* $15,000 *per month per stock."* in "Firms analyze tweets to gauge market sentiment," WSJ, July 6, 2015.
- **Demand :** *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.
- **Many information sellers :** Thomson Reuters, Bloomberg, Ravenpack, Dataminr, Eagle alpha, iSentium, Thinknum, Psychsignal, TheySay, MarketPsych, MarketProphit, Orbital, Cargometrics etc.

Data Abundance

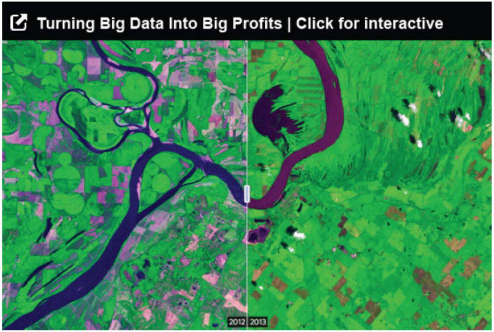


FIGURE: source : Orbital Insight/WSJ 2014

Data Abundance : Curse or Blessing ?

- **Does a decline in the cost of access to information make asset prices more informative ?**

Data Abundance : Curse or Blessing ?

- **Does a decline in the cost of access to information make asset prices more informative ?**
- **Existing models of information acquisition :** If the cost of information declines :
  1. More investors buy information (Grossman and Stiglitz (1980)) or investors buy more precise signals (Verrechia (1982))
  2. $\Rightarrow$ Asset price informativeness increases.

Data Abundance : Curse or Blessing ?

- **Does a decline in the cost of access to information make asset prices more informative ?**
- **Existing models of information acquisition :** If the cost of information declines :
  1. More investors buy information (Grossman and Stiglitz (1980)) or investors buy more precise signals (Verrechia (1982))
  2. $\Rightarrow$ Asset price informativeness increases.
- However, information processing is instantaneous in these models : no lag between getting data and processing the data.

Data Abundance : Curse or Blessing ?

- **Does a decline in the cost of access to information make asset prices more informative ?**
- **Existing models of information acquisition :** If the cost of information declines :
  1. More investors buy information (Grossman and Stiglitz (1980)) or investors buy more precise signals (Verrechia (1982))
  2. $\Rightarrow$ Asset price informativeness increases.
- However, information processing is instantaneous in these models : no lag between getting data and processing the data.
- **In reality :**
  1. More data $\neq$ More accurate signals
  2. Big (raw) data are very noisy
  3. Filtering out noise from data takes time.
  4. $\Rightarrow$ **Trade-off between trading early on very noisy signals or later on accurate signals.**

Noisy signals ? 'A rose by any other name'...

- On January 14, 2014 : Google announced its deal to acquire a private firm Nest Labs. NEST stock (Nestor Inc) is not Nest lab...and Nestor Inc is bankrupt...

Noisy signals ? 'A rose by any other name'...

- On January 14, 2014 : Google announced its deal to acquire a private firm Nest Labs. NEST stock (Nestor Inc) is not Nest lab...and Nestor Inc is bankrupt...

Data Abundance and Noise : The Twitter Crash

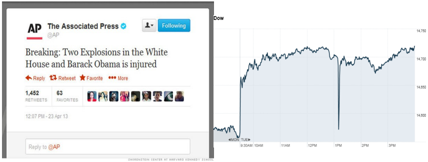


FIGURE: The "Twitter Crash" of April 2013

"Twitter's credibility (a novel idea to non-tweeters) has taken a hit. But 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." (**Source :** The Economist, April 27 2013).

Why should we care ?

- An important function of financial markets is to **produce new information** that can be used by real decision makers (see Edmans, Bond and Goldstein (2012) for a survey).
  1. **Fama and Miller (1972, p.335) :**"*An efficient market has a very desirable feature. In particular, at any point in time market prices of securities provide accurate signals for resource allocation ; that is firms can make production-investment decisions.*"

Why should we care ?

- An important function of financial markets is to **produce new information** that can be used by real decision makers (see Edmans, Bond and Goldstein (2012) for a survey).
  1. **Fama and Miller (1972, p.335) :**"*An efficient market has a very desirable feature. In particular, at any point in time market prices of securities provide accurate signals for resource allocation ; that is firms can make production-investment decisions.*"
  2. **Important for policy : Is lowering access cost to raw data (e.g., through on-line access of accounting information) a good idea ?**

Why should we care ?

- An important function of financial markets is to **produce new information** that can be used by real decision makers (see Edmans, Bond and Goldstein (2012) for a survey).
  1. **Fama and Miller (1972, p.335) :**"*An efficient market has a very desirable feature. In particular, at any point in time market prices of securities provide accurate signals for resource allocation ; that is firms can make production-investment decisions.*"
  2. **Important for policy : Is lowering access cost to raw data (e.g., through on-line access of accounting information) a good idea ?** "*If [XBLR] serves to lower the data aggregation costs [...] smaller investors will have greater access to financial data than before. [...]. Lower data aggregation costs will allow investors to either aggregate the data on their own, or purchase it at a lower cost [...]. Hence, smaller investors will have fewer informational barriers that separate them from larger investors with greater financial resources.*" (SEC (2009))

Evidence

- **Long-term trend in price informativeness is unclear.** Bai, Phillipon, and Savov (2014) find a decline in price informativeness for the entire universe of U.S stocks (an increase for $S\&P500$ stocks) .
- **Effects of algorithmic trading on price informativeness is unclear.**
  1. Makes prices more efficient (fewer arbitrage opportunities or prices closer to random walks ; see Chaboud et al.(2014) or Brogaard et al.(2015)
  2. But does this make prices more informative ? (Weller (2016) finds a negative association between algo trading and price informativeness).
- Does more data make prices more informative at the horizon relevant for real decision makers ?

Our Model

- **Speculators can buy two types of signals :**
    1. **Unfiltered signals** = Information or Noise ("News-Analytics")
    2. **Filtered signals** = without noise ("Financial Analysis")
- **Filtered signals can only be obtained with a lag relative to unfiltered signals.**
- **The prices of both signals are endogenous, set by competitive information sellers** (e.g., Thomson, Dataminr, iSentium and "financial analysts").
- **We solve for equilibrium strategies (demand for unfiltered and filtered information)**, **prices and trades and analyze the effect of data abundance (a reduction in the cost of accessing unfiltered information) on (i) asset price informativenes and (ii) lead-lag relationships between prices and trades.**

Findings

- **A decline in the cost of accessing unfiltered information :**

1. Increase the demand for unfiltered information ⇒ make prices more informative in the short run.

2. Can reduce the expected profit of trading on filtered information ⇒ Reduce the demand for filtered information ⇒ Reduce long run price informativeness.

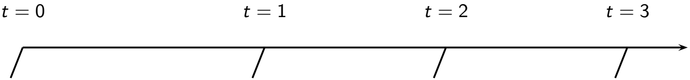
3. Increase the absolute value of the correlation between trades of investors trading on filtered signals with past returns (correlation can be positive or negative depending on signals quality).

4. Reduce the correlation in trades of investors trading on unfiltered and filtered signals and can even induce a negative correlation in these trades

Literature

- **Costly information acquisition and markets for information** (e.g., Grossman and Stiglitz (1980), Verrechia (1982), Admati and Pfleiderer (1986), Veldkamp (2006), Lee (2013)).
    1. In some models, investors can acquire more precise information (e.g., Verrechia (1982)) at a cost.
    2. But information is instantaneously available.
- **Early and late informed traders :** Froot, Scharfstein and Stein (1992), Hirshleifer, Subrahmanyam, and Titman (1994), and Brunnermeier (2005). Differences :
    1. No endogenous choice to trade late or early in these papers.
    2. The precision of signals for late and early traders is the same.
    3. $\Rightarrow$ predictions are different (e.g., the predictions about relationships between returns and trades are different)

Model



$t = 0$

Markets for information :

- A mass α_1 of speculators decide to buy the raw signal, which will be available at date 1, at price F_r.
- A mass α_2 of speculators decide to buy the processed signal, which will be available at date 2, at price F_p.

$t = 1$

- Speculators observe the raw signal s, then submit buy or sell orders for one share.
- Liquidity traders submit buy or sell orders.
- The market maker observes the aggregate order flow, f_1, and sets a price p_1.

$t = 2$

- Speculators observe the processed signal (s, u), then they submit buy or sell orders for one share.
- Liquidity traders submit buy or sell orders.
- The market maker observes the aggregate order flow, f_2, and sets a price p_2.

$t = 3$

The asset pays off, $V \in \{0, 1\}$.

Modeling Information Processing

- **A continuum of speculators : At date** 0, each speculator can buy two different signals :
  1. **A "raw" (unfiltered) signal at price (fee) F_r :**

     $$S = U \times \underbrace{V}_{fundamental} + (1 - U) \times \underbrace{\epsilon}_{Noise},$$

     where $U = 1$ or 0 with prob. θ ; $\epsilon = 1$ or 0 with prob. $1/2$; and $\epsilon \perp V$.
  2. **A "processed" (filtered) signal at price F_p :** i.e., a signal (S, U).

- $\theta =$ "Raw signal reliability"
- **Assumption :** Filtering out noise from signals takes time $\Rightarrow$ Processed information is available with a lag of one trading period relative to the raw signal.
- We denote by α_t the mass of speculators buying the signal available at date t (= demand for information).

Trading

- **3 types of market participants at dates 1 and 2 :**
  1. **Liquidity Traders.** Their aggregate trade at date t, l_t is uniformly distributed on $[-1, 1]$.

  2. **Speculators :** Optimally decide to buy/sell one share of the asset at dates 1 or 2 after observing their signal.
     - 2.1 **Date 1 : Speculators trading on the raw signal :** Collectively buy or sell α_1 shares.
     - 2.2 **Date 2 : Speculators trading on the processed signal :** Collectively buy or sell α_2 shares.
  3. **Market makers.** Risk neutral and competitive. They absorb the aggregate order imbalance (net demand of liquidity traders + speculators) at price :

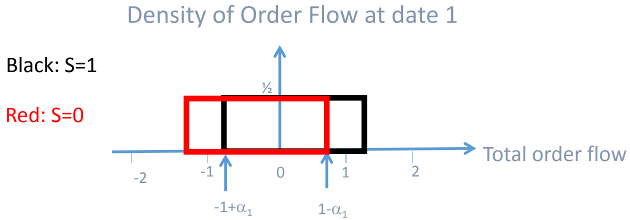
$$p_t = E(V \mid \Omega_t),$$

where Ω_t = History of order imbalances until date t.

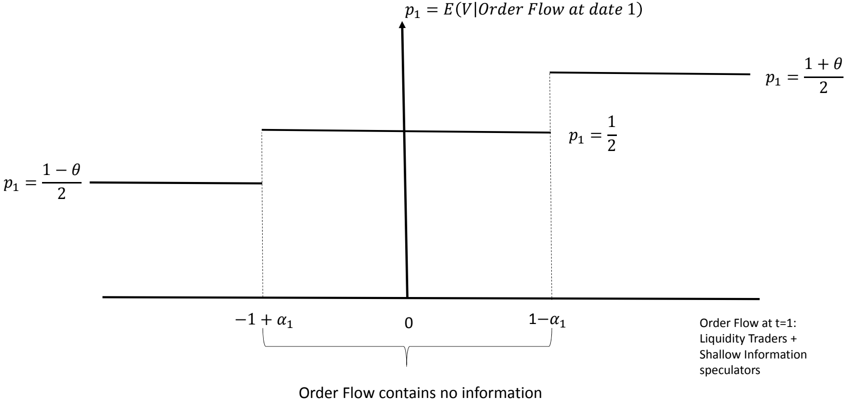
Next Steps

- Equilibrium prices and trading strategies at dates 1 and 2 for given demands for raw and deep information.
- Equilibrium prices of and demands for deep and raw information at date 0
- Implications.

Equilibrium distribution of order flow at date 1



Equilibrium price at date 1 (standard)



Speculators' Expected Profits at date 1

- **Gross expected profit of a raw information speculator :**

$$\pi_1(\alpha_1) = \frac{\theta}{2} \times (1 - \alpha_1) .$$

- Increases in signal reliability, θ and decreases in the mass of raw information speculators, α_1.
- Likelihood that raw information speculators gets reflected into prices at the end of date 1 : α_1.
- "Maximal capacity" of the "buy raw information strategy :" $\alpha_1 = 1$.

Trading on the Processed Signal

- **Case 1 : The price at date 1 reflects the raw signal ($p_1 = s \neq p_0$)**

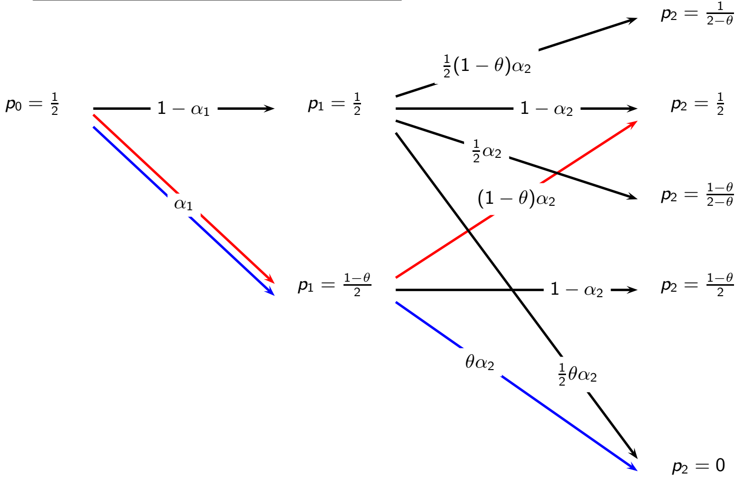
  1. If the raw signal is noise, speculators with the processed signal correct the noise in price : they trade in a direction opposite to past returns (from date 0 to 1).
  2. If the raw signal is not noise, speculators with the processed signal trade on the fundamental : they trade in the same direction as past returns (from date 0 to 1).

- **Case 2 : The price at date 1 is uninformative ($p_1 = p_0$)**

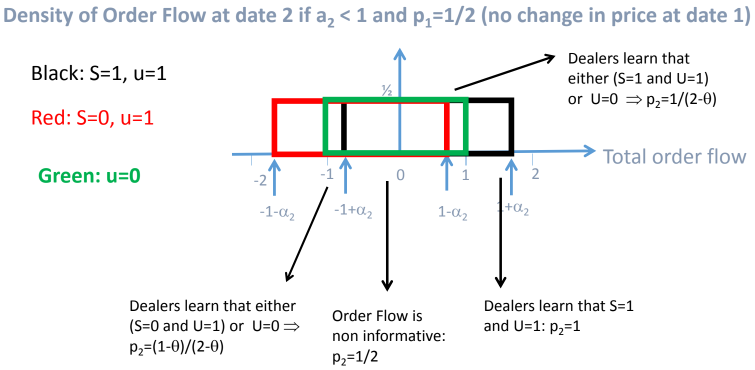
  1. Speculators with the processed signal can only trade profitably if the raw signal is not noise.

Price Dynamics 2/2

Price dynamics conditional on $s = 0$



Equilibrium distribution of order flow at date 1 if $p_1 = 1/2$



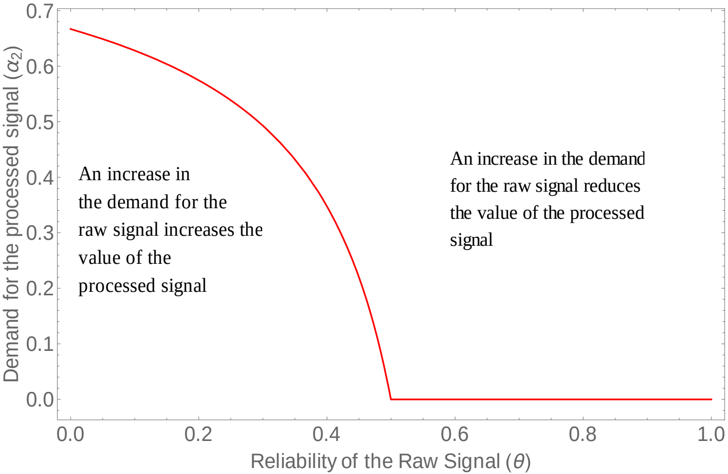
The Value of Processing Information

- **The ex-ante expected profit from trading on the processed signal ($\bar{\pi}_2(\alpha_1, \alpha_2)$) is :**

$$\alpha_1 \times E(Profit(\alpha_2) \mid p_1 = s) + (1 - \alpha_1) \times E(Profit(\alpha_2) \mid p_1 = p_0) \quad (1)$$

- **Standard :** An increase in the mass of speculators trading on the processed signal (α_2) reduces the return from trading on this signal.
- **Not Standard :** An increase in the mass of speculators trading on the raw signal (α_1) can reduce or increase the return from trading on this signal (depends on θ and α_2).

When does demand for the raw signal degrade the value of the processed signal?



The Market for Information

- **Producing a given type of signal : fixed cost but zero marginal cost** (as in Veldkamp (2006)).
  1. C_p the fixed cost of producing the processed signal.
  2. C_r the fixed cost of producing the raw signal.

- **Markets for information are competitive :**
  1. Fees for each type of signal adjust so that sellers just cover the fixed cost of producing a signal $\Rightarrow$

$$Price \quad of \quad Signal = \frac{\text{Fixed Cost of Producing the Signal}}{\text{Number of Buyers}}.$$

  2. Entry of new speculators until expected profits on information net of price equal zero. $\Rightarrow$

$$Speculators \quad Aggregate \quad Profits = Fixed \quad Cost \quad of \quad Information.$$

Equilibrium in the Market for the Processed Signal 1/2

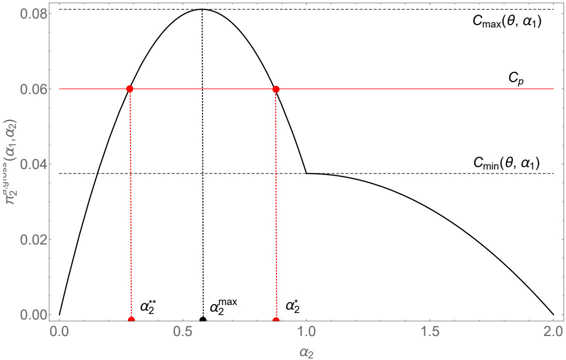


FIGURE: Note : In equilibrium the aggregate gross profit from trading on the processed signal $\alpha_2 \times \bar{\pi}_2^{gross}$ is equal to the cost of producing this signal.

Equilibrium in the Market for the Processed Signal 2/2

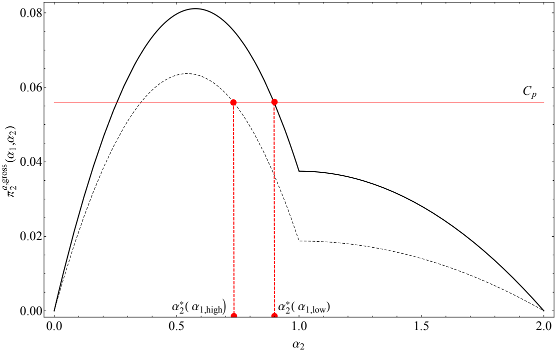


FIGURE: If $\theta > \frac{\sqrt{2}-1}{\sqrt{2}}$, a decrease in the cost of producing the raw signal reduces the equilibrium demand for the processed signal.

Crowding Out

- **Result :** When $\theta > \frac{\sqrt{2}-1}{\sqrt{2}}$, there exist parameter values such that the demand for deep information drops to zero when the cost of raw information goes to zero.

Crowding out Processed Signals Speculators

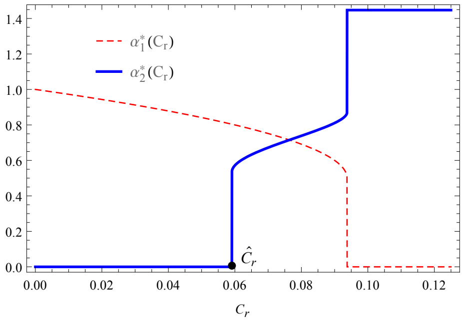


FIGURE: X-Axis : Cost of Producing the Raw Signal. RED : Demand for the Raw Signal. BLUE : Demand for the Processed Signal.

Implications : Asset Price Informativeness

- **Price informativeness at date t :**

$$\mathcal{E}_t(C_r, C_p) = \frac{1}{4} - E[(\tilde{V} - P_t)^2] = \frac{1}{4} - E[Var[V|\Omega_t]].$$

- **Does a reduction in the cost of the raw signal make prices more informative ?**
  1. In the short run ? (does $\mathcal{E}_1(C_r, C_p)$ decrease with C_r ?)
  2. In the long run ? (does $\mathcal{E}_2(C_r, C_p)$ decrease with C_r ?).

Remarks

- **Remark 1 :**
  1. A reduction in the cost of producing raw information increases the demand for raw information and therefore makes prices more informative in the short run.
  2. A reduction in the cost of producing deep information makes prices more informative in the long run.
  3. Not surprising : standard in models of trading with endogenous information acquisition (e.g., Grossman and Stiglitz (1980)).

- **Remark 2 :**
  1. Prices are necessarily more informative at date 2 (the long run) than at date 1 because information accumulates over time ($\Omega_1 \subseteq \Omega_2$) :
  $$\mathcal{E}_2(C_r, C_p) \geq \mathcal{E}_1(C_r, C_p).$$
  2. Yet, we might have : $\mathcal{E}_2(C_r, C_p)$ decreases while $\mathcal{E}_1(C_r, C_p)$ increases when C_r decreases.

Long run asset price informativeness and data abundance

- **Result :** A reduction in the cost of producing the raw signal can reduce asset price informativeness in the long run.
- **Intuition :**
  1. Reduction in the cost of the raw signal $\rightarrow$ Increase in demand for the raw signal.
  2. $\rightarrow$ Expected return from trading on the processed signal declines.
  3. $\rightarrow$ Demand for the processed signal declines $\rightarrow$ Long run asset price informativeness drops.

Example

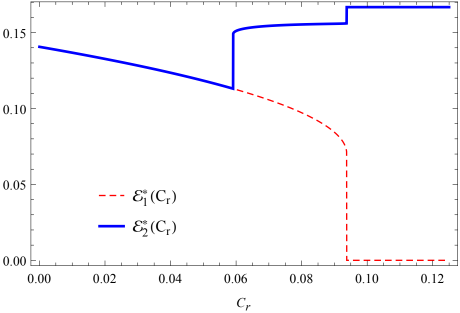


FIGURE: X-axis : Cost of producing the raw signal. BLUE : Asset price informativeness at date 1 RED : Asset price informativeness at date 2.

Joint decrease in the cost of producing raw and shallow information ?

- Progress in information technologies should reduce both the cost of producing shallow information (C_r) and the cost of processing information (C_p).
- Yet, this might not be sufficient if C_r declines faster : for any level of C_p, one can find a level of C_r small enough such that long run price informativeness is less than if trading on shallow information was impossible.

Price and Trade Patterns

- **Suppose you have data on trades by speculators trading on processed signals ("deep information speculators") and speculators trading on the raw signals ("raw information speculators").**
- **What are the effects of a decrease in the cost of producing the raw signal on :**
  1. The relationship (covariance) between order flows of both types of speculators ?
  2. The relationship between the order flow of speculators who trade on processed signals and past returns ?
  3. The relationship between the order flow of speculators who trade on raw signals and future returns ?

Raw and Deep Information Speculators' Order Flows

- **Result :** The covariance between raw and deep information speculators' orders ($Cov(x_1, x_2)$)

  1. Is (i) **positive** for $\theta > 1/2$ and can be **negative** if $\theta < 1/2$ and $C_r < \bar{C}_r(\theta)$.

  2. Should become smaller when the cost of raw information declines.

- **Prediction :** Data abundance reduces the correlation between raw and deep information speculators' orders.

Example

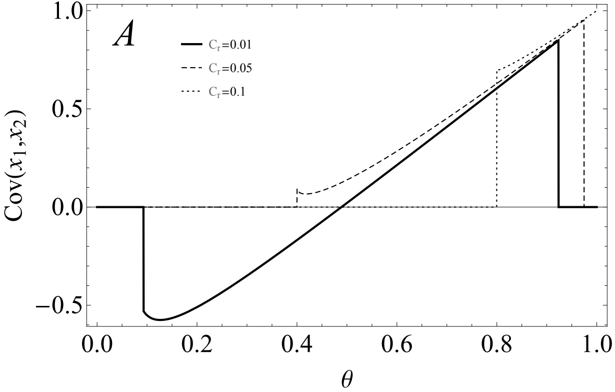


FIGURE: Y-Axis : $Cov(x_1, x_2)$. X-Axis : θ.

Returns and Trades 1/2

- **Result :** The covariance between deep information speculators' orders and past returns ($Cov(p_1 - p_0, x_2)$)
  1. Is (i) **positive** if $\theta > 1/2$ and **negative** if $\theta < 1/2$.
  2. Should become larger in absolute value when the cost of raw information declines.
- **Prediction :** Data abundance increases the absolute value of the correlation between deep information speculators' trades and past returns.
- Deep information speculators can appear (to the econometrician) following either a *momentum* strategy or a *contrarian* strategy.

Example

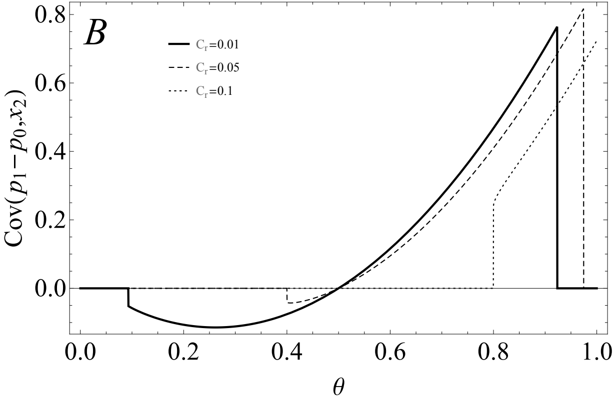


FIGURE: Y-Axis : $Cov(p_1 - p_0, x_2)$. X-Axis : θ.

Returns and Trades 2/2

- **Result :** The covariance between raw information speculators' orders and future returns ($Cov(x_1, p_2 - p_1)$)
  1. Is **positive** (raw information speculators trade on information...).
  2. Should become smaller when the cost of raw information declines.
- **Prediction :** Data abundance reduces the correlation between raw information speculators' trades and future returns

Example

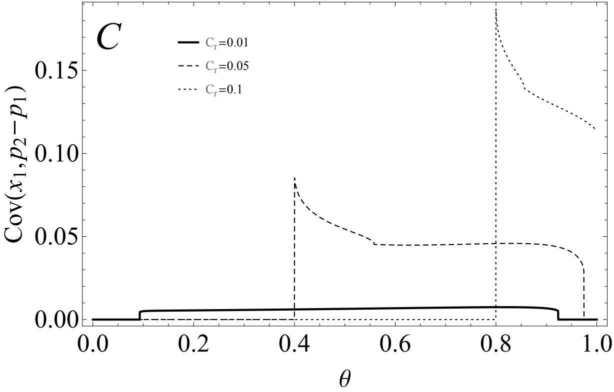


FIGURE: Y-Axis : $Cov(x_1, p_2 - p_1)$. X-Axis : θ.

Conclusions

- **Data abundance can reduce asset price informativeness.**

Conclusions

- **Data abundance can reduce asset price informativeness.**
- **Next steps : Empirical tests**

1. **Find exogenous shocks to the cost of access to financial information** (e.g., digitalization of accounting information by firms) and their effects on (i) the production of information (e.g., number of financial analysts per firm) + (ii) price informativeness + (iii) real decisions (firms' investment).

2. **Check whether patterns of prices and trades for various groups of investors fit the predictions of the model.**

3. **More detailed analysis of the competition between raw information sellers (Bloomberg, Thomson-Reuters etc.) and deep information producers (financial analysts).**