WHERE HAS ALL THE BIG DATA GONE?

Maryam Farboodi
Princeton

Adrien Matray
Princeton

Laura Veldkamp
NYU Stern School of Business

2018
Motivation

- Increase in “big data” in financial sector
  1. ↑ data processing power
  2. ↑ resources invested by financial firms

- Q: Has data-related technology adoption improved financial efficiency?
  - One approach to measuring financial efficiency: Price informativeness
    Bai, Philippon and Savov (2016) say: Price info increased.
  - We find: Price informativeness increases only for S&P 500 firms. This is a size composition effect.
  - For non-S&P 500 firms, price information declines.

- Do these facts imply that big data is a big flop?
  We use a model to show: not necessarily.
  But the benefits of big data may be accruing to a small set of large firms.
Data

- Stock prices: CRSP 1962-2015
- Accounting variables: Compustat
- Institutional holdings: 13-f
Defining Price Informativeness

- Price ability to predict future cash flow: Follow Bai, Philippon & Savov (‘16)

\[
\frac{Ebit_{j,t+k}}{Asset_{j,t}} = \alpha + \beta_t \log \left( \frac{MkVal_{j,t}}{Asset_{j,t}} \right) + \gamma X_{j,t} + \epsilon_{i,t} \quad \text{with } k = [3, 5]
\]

\[
\left( P_{Info} \right)_t = \beta_t \times \sigma_t \left( \log(M/A) \right)
\]

- Effect of technology adoption by financial sector?
**Price Informativeness Diverges**

**S&P 500 Price Info Increases**
(BPS result)

**Non-S&P 500 Price Info Decreases**

**Our question:** What does this mean? Does the decline overall disprove the BPS hypothesis that information processing advances are increasing financial efficiency?
**Key Fact: Large Firms are Getting Larger**

### Market Value (in $M)

<table>
<thead>
<tr>
<th>Decade</th>
<th>Average Size (M. Value $M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960-1970</td>
<td>49.7</td>
</tr>
<tr>
<td>1970-1980</td>
<td>54.7</td>
</tr>
<tr>
<td>1980-1990</td>
<td>54.7</td>
</tr>
<tr>
<td>1990-2000</td>
<td>147.8</td>
</tr>
<tr>
<td>2000-2010</td>
<td>228.8</td>
</tr>
</tbody>
</table>

### Share Firms in Size Decile 9

- **In S&P500**: Blue bars
- **Not in S&P500**: Orange bars

![Graph of Share Firms in Size Decile 9](image-url)
Larger Firms’ Prices More Informative

\[
\frac{E_{i,y,t+k}}{A_{i,y,t}} = \alpha + \beta_{t,y} \log \left( \frac{M_{i,y,t}}{A_{i,y,t}} \right) + \gamma X_{i,y,t} + \epsilon_{i,y,t}
\]

with \( y \in [1, \ldots, 10] \) bin size
Size Growth Explains Increase in Price Informativeness

Size-predicted Price Info: $\hat{beta}_t = \sum_{y \in [1, \ldots, 10]} \beta_y \times ShareFirms_{y,t}$

If size alone can explain the rise in S&P 500 price info, . . . that doesn’t look promising for BPS’ financial efficiency story.
... BUT NOT THE Decrease in Price Informativeness

Overall PI should have increased

As small firms increased

- To make matters worse, decrease in price info is not explained.
- Perhaps small firms got more tech-y? In paper: they did, but only explains small part of decline.
- Did financial markets get less efficient?
But growth firms in S&P 500 account for most of the increase too! No composition explanation (wrong direction). Just a clue.

All publicly listed, nonfinancial firms in given year. Growth firms = bottom 30% B/M; value = top 30%. PI from a separate regression for each year \( t = 1962, \ldots, 2010 \).
Forces that Do Not Explain These Trends (in the paper)

- **Membership in index** is not key. Similar *upward* trend in price informativeness for:
  - Former and future members
  - Firms similar to S&P 500 firms but **never** members

- **Industry** is not key:
  - Few industries represented in S&P 500 (only 24 SIC3-industries with 12 firms or more in S&P 500 ≠ 253 in whole sample)
  - But same *divergence* in price informativeness *within* these 24 industries

- **Tech** firms or high R& D firms. These are harder to value (lower PI). There are more small tech firms. But this shifts are too small to explain much.
If firms aren’t making better investments, why spend all this? Is there some part of fin tech being used to learn more about assets?
Open Questions

- A very different picture of information in prices from Bai et. al.

- What do these facts imply? Do they prove that financial markets have become less efficient data processors over time?

- To know if data disprove the efficiency story, we need a model with an increase in efficiency so we can see if the data rejects it.

- The model needs to have:
  - Choice of what data to process
  - Investment choice
  - Different size firms
CARA normal model with 1 riskless, \( n \) risky assets and stochastic supply

- Risky asset \( \equiv \) claim to firm dividend
  
  - Dividend: \( \tilde{d}_t = g(l_t) + \tilde{y}_t \) \( \tilde{y}_t \sim \mathcal{N}(0, \Sigma) \) \[ \text{no } j: \text{vector} \]
  
  - Stochastic supply: \( \bar{x} + \tilde{x}_t \) \( \tilde{x}_t \sim \mathcal{N}(0, \Sigma_x) \) \[ \sim: \text{unknown at } t \]

- A continuum of 1-period-lived investors, demand \( q_{it} \) risky assets

- Market clears: \( \int q_{it} \, di = \bar{x} + \tilde{x}_t \)
At $t$, investors use priors, information from prices, and data they choose to see, to form beliefs.

**Data**: a binary code (bits) about dividend. A string of binary code with Gaussian noise is like observing a signal:

$$\eta_{jt} = \tilde{d}_t + \tilde{\epsilon}_{jt} \quad \tilde{\epsilon}_{jt} \sim \mathcal{N}(0, K_{jt}^{-1})$$

where bits:

$$\text{bits} = \frac{1}{2} \log(1 + \text{signal-to-noise})$$

⇒ To get precision $K_{jt}$ about asset $j$, investor must allocate

$$b_j = \frac{1}{2} \log(1 + K_{jt} / \Sigma_{jj}^{-1})$$

**Technological progress**: Growth in total bits processed

⇒ Across which assets investors allocate this increasing amount of data processing power?
Optimal mean variance portfolio $q_t$ given information set

Market clearing price:

$$p_t = A + C\tilde{y}_t + D\tilde{x}_t \quad C : \text{Price Informativeness}$$

Investors’ data choice problem: For which asset do investors process data? Depends on marginal value of data precision ($K_{jt}$). Two key forces:

- **Strategic substitutability** in data processing
- **Returns to asset scale** in data processing
**Equilibrium Data Choice**

**Lemma**

Technological progress makes all price informativeness grow. As total data processed grows, the average investor learns weakly more about every asset $j$, $\int K_{ij(t+1)} dt \geq \int K_{ijt} dt$, with strict inequality for assets that are learned about.

**Lemma**

Relative-size growth explains the divergence in price informativeness. For sufficiently high amount of total data processed, an increase in the size of firm $j$ increases the amount learned about $j$ and reduces the amount learned about all other assets: $\partial \bar{K}_{jt} / \partial \bar{x}_j > 0$ and $\partial K_l / \partial \bar{x}_j \leq 0 \forall l \neq j$.

- Key to different trends is size divergence
Price Informativeness. Relative Size and Data Technology Grow

- Model can explain the empirical divergence in small and large firm price informativeness
What About The Growth Firms?

- Long-lived assets have values that depend on $\tilde{d}_t/(r - G)$. The higher is $G$ (growth), the more sensitive asset values are to news ($\tilde{d}_t$).

- If a small change in $\tilde{d}_t \rightarrow$ big change in value, news about $\tilde{d}_t$ more valuable ($dK_j/dG_j > 0$).
  → More data about growth firms.

- What does this mean for observed trends?
  - Suppose investors were mostly learning about growth firms.
    (This requires $G$ for growth firms to be sufficiently high.)
  - Then when large firms get larger, the large growth firms attract more data and the small growth firms’ data gets crowded out.
Did financial efficiency increase?

- Yes: for S&P 500 firms, but this is a composition effect.
- No: For small firms, efficiency fell

Does this disprove the hypothesis that big data was adopted and used optimally? No.

- Big firms grew larger. Larger firms have more valuable data.
- This can draw data away from small firms
- The facts are consistent with Bai et. al.’s conjecture that more information processing was taking place. But the facts and the logic are more subtle than we knew.
Price Informativeness Trend While in and out of S&P 500 is Similar

The grey line (bottom) is the firms currently in the S&P 500, at the date listed on the x-axis. The red line (top) is firms not currently in the S&P 500. The black and red dashed lines are linear trends that fit the grey and red time trends, respectively.
We run a separate regression for each year $t = 1962, \ldots, 2010$, horizon $k = 5$ and bin interval $[1/10), \ldots [10/10]$ partitioned by 1/10 deciles of firm average R&D spending, scaled by its assets. All publicly listed nonfinancial firms from 1962 to 2010.
Predicted PI based on High-Tech: S&P 500 and Whole Sample

Predicted PI = \sum_{y\in[1,\ldots,10]} \beta_y \times ShareFirms_{y,t}.
Growth firms are bottom 30% of $B/M$; value firms are top 30%.
S&P 500 Growth Firms Grew Bigger
SMALL AND LARGE FIRM RETURNS ARE NOT DIVERGING
**Numerical Example**

- Risk aversion is $\rho = 4$
- The inverse variance of the dividend payoff is $\Sigma^{-1} = 1$
- The inverse variance of asset supply shock is $\tau_x = 3$
- Firm size: $\bar{x}_1$ and $\bar{x}_2$ both start at 1. But $\bar{x}_1$ grows by 0.1 each period. The small firm size stays constant.
- The total data processing capacity grows at a constant rate. Starts at $K_t = 8$ and ends at $K_t = 16$. 