Where Has All the Big Data Gone?

Maryam Farboodi* Adrien Matray† Laura Veldkamp‡

July 6, 2017§

Abstract

As the size of the financial sector has increased and ever more technology is deployed to process and transmit financial data, this could benefit society by allowing capital to be allocated more efficiently. Recent work supports this notion. Bai, Philippon, and Savov (2013) document an improvement in the ability of financial prices to predict firms' future earnings. This "price informativeness" rises for firms in the S&P 500 . We show that most of this rise comes from a composition effect. S&P 500 firms are getting older and larger. In contrast, the ability of the market to price new, small and growing firms, from whom most of productivity growth comes, is deteriorating. To understand the causes and social consequences of this shift, we formulate a model designed to show why investors prefer to process large firm data. We then use the model to explore why individual investors might make data processing decisions that deviate from the social optimum. The model provides a possible explanation for why the ever-growing reams of data processed by the financial sector have not delivered tangible, real, economic benefits, for the vast majority of firms.

Does the growth in the financial sector add social value? The answer to this basic question lies at the heart of many policy and regulatory debates. Recent evidence that the informativeness of asset prices has been increasing (Bai, Philippon, and Savov, 2013) suggests that the labor, technology, and human capital growth in the financial sector is yielding real benefits, in terms of more efficient capital allocation for firms. However, this rosy headline result of greater price informativeness pertains to firms in the S&P 500. More accurately,

*Princeton University; farboodi@princeton.edu

†Princeton University;amatray@princeton.edu

‡Department of Economics Stern School of Business, NBER, and CEPR, New York University, 44 W. 4th Street, New York, NY 10012; lveldkam@stern.nyu.edu; http://www.stern.nyu.edu/∼lveldkam.

§Preliminary and incomplete. We thank John Barry, Matias Covarrubias for their capable research assistance. JEL codes: Keywords: financial technology, big data, capital misallocation.

large, old firms are priced, and have always been priced, more efficiently. These old firms have simply become more prevalent in the S&P 500. For the universe of publicly traded firms as a whole, price informativeness has, in fact, declined. This paper explores these competing facts, teases out composition effects from trends, explores the reasons for the shifts in market efficiency, and concludes that the growth of data processing may be creating net losses for the real economy.

Of course, S&P 500 firms constitute 70-80% of total market capitalization. From a financial perspective, these are the most relevant firms. But these are not the fastest-growing, not the highest-productivity, and not the most job-creating firms. From a social perspective, allocating capital properly to non-S&P 500 firms might be more important. This possibility motivates our exploration of what kinds of firms are being priced less informatively and why.

Section 1 starts by exploring the question: What is it about S&P 500 firms that explains why their prices became more informative when other firms' price informativeness has fallen? Is current membership in the S&P important? No, we compare the set of all firms that have been in the S&P 500 at some time, and find that price of firms currently in the S&P 500 have neither a higher level or steeper trend in price informativeness. Is this is an industry effect? No, we find that for industries most represented in the S&P, the firms in those industries that are not themselves S&P 500 members, those firms' prices have witnessed no rise in price information. Next, we explore firm age. This shows some promise. Older firms do indeed have more informative prices. However, the informativeness of the oldest firms in the same has not risen. Since S&P 500 firms are undoubtedly older, and older firms have more informative prices, this raises the possibility that of an age composition effect. We find that S&P 500 firms are getting older over time. This shift in average firm age can account for most of the rise in the informativeness of S&P 500 prices.

The change in size composition of S&P 500 firms explains the growth of price informativeness of the S&P 500 . This finding leaves two open questions. First, why does the price informativeness of the non-S&P 500 firms decline? Price informativeness is not a zero-sum game. Information processing is growing and computing power grows. It is perfectly possible that price informativeness of S&P 500 firms grows without causing a decline in the informativeness of other firms' prices. Second, what are the real economic consequences of

this trend?

To answer these two questions, Section 3 develops a model with a few key ingredients. First, we need an asset market environment, with multiple assets, and where investors may choose to process data about any or all of those assets, at a cost. Second, for data processing to affect prices, investors must use their data to make portfolio investment choices and the price must clear the market. Third, asset market outcomes must affect real output and real efficiency.

The model teaches us why processing large firm data is more valuable to an investor than processing small firm data. However, as total data processing rises, results reveal that investors should not only examine additional data on large firms. A robust prediction of the model is that data processing for all firms should rise. If increasing data processing is the only force at work, the firms' price informativeness should rise across the board. The fact that large firm informativeness rises and small firms' informativeness stagnates or falls means that some other force, besides more data, must be at work.

Next, we use the model to ask whether changes in the firm size distribution could explain the changes in price informativeness. We find that if the largest firms grow larger, this can explain why large firm prices become more informative and small firm prices less informative. The large firms become more attractive targets for data processing and they draw attention away from the relatively less attractive small firms.

These findings are consistent with a financial sector that has or has not improved its ability to process data and use that knowledge to price assets. The rise in S&P 500 price informativeness does not necessarily mean that financial analysis is better. But these facts do allow us to bound the increase in data productivity. If the rise in productivity is too large, relative to the increase in the size of large firms, then such a combination of forces would be unable to explain the stagnation in price informativeness of small firms.

Finally, we explore welfare. It is efficient for investors to process more data on large firms than small ones, because large firm outcomes matter more for aggregate output. However, investors do over-process large firm data, relative to the social optimum. The reason is that data processing leads investors to want to hold more of the studied asset, on average, which further raises the value of data. In an economy, the shares of each firm are given. So the

increasing returns to size that come from a data-portfolio choice feedback are not present at an aggregate level. If in the future, all data economy-wide could be processed in an integrated way, instead of parallel processing, then the social optimum would shift, favoring more processing of large firm data.

Our Contribution Relative to the Literature Many papers have developed approaches to measuring stock market informativeness across countries (Edmans, Jayaraman, and Schneemeier, 2016), or (Durnev, Morck, and Yeung, 2004). The novelty of our approach, compared to these studies is that we are interested in how price informativeness evolves over time because it reveals changes in the efficiency of the financial sector.

Examinations of the effects of big data are scarce. Empirical work primarily examines whether particular data sources, such as social media text, predict asset price movements (Ranco, Aleksovski, Caldarelli, Grcar, and Mozetic, 2015). On the theoretical side, the information theory (computer science) based measures we use to quantify big data flows are similar to those used in work on rational inattention (Sims, 2003; Maćkowiak and Wiederholt, 2009; Mondria, 2010; Kacperczyk, Nosal, and Stevens, 2015). While the types of information choices made in these models is similar, most are about macroeconomic questions or about cross-sectional, rather the long-run phenomena.

Our work also contributes to the debate on the sources of capital misallocation in the macroeconomy (see Restuccia and Rogerson (2013) for a survey). Most work in this area focuses on how efficiently, or inefficiently, capital is allocated (e.g. Hsieh and Klenow (2009)). Like David, Hopenhayn, and Venkateswaran (2016), our focus is on the role financial markets play in informing these allocation or investment choices. What we add is a deeper understanding of why financial markets may be providing better guidance over time for some firms, but not for others.

Starting with Grossman and Stiglitz (1980), equilibrium models with information choice have been used to explain income inequality (Kacperczyk, Nosal, and Stevens, 2015), home bias (Mondria, Wu, and Zhang, 2010; Van Nieuwerburgh and Veldkamp, 2009), mutual fund returns (Pástor and Stambaugh, 2012; Stambaugh, 2014; Kacperczyk, Van Nieuwerburgh, and Veldkamp, 2016), growth of financial sector (Farboodi and Veldkamp, 2017), among

other phenomena. Related microstructure work explore the frequency of information acquisition and trading (Kyle and Lee, 2017; Dugast and Foucault, 2016; Chordia, Green, and Kottimukkalur, 2016; Crouzet, Dew-Becker, and Nathanson, 2016). Davila and Parlatore (2016) share our focus on price information, but do not examine its time trend or cross-sectional differences.

Our model extends Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) in three ways. First, our information processing constraint corresponds to computer science measures of data processing, based on bits. That change allows us to map processing directly to CPU speed. Second, we have heterogeneous asset characteristics, which allow us to explain how data processing and thus price informativeness should covary with asset characteristics, such as size. Finally, we add a link between the financial and real economies that allows us to address long-run changes in aggregate output and welfare.

1 Data and Measurement of Price Informativeness

1.1 Data

The data we are using are for the U.S. over the period 1964–2015. Stock prices come from CRSP (Center for Research in Security Prices). All accounting variables are from Compustat. We take stock prices as of the end of March and accounting variables as of the end of the previous fiscal year, typically December. This timing convention ensures that market participants have access to the accounting variables that we use as controls.

The main equity valuation measure is the log of market capitalization M over total assets A, $log(M/A)$ and the main cash flow variable is earnings measured as EBIT (earnings before interest and taxes, denoted $EBIT$ in Compustat). This measure includes current and future cash flows, and investment by current total assets. All ratio variables are winsorized at the 1%.

Since we are interested in how well prices forecast future earnings, and future earnings are affected by inflation, we need to consider how to treat inflation. We adjust for inflation with GDP deflator to ensure that differences in future nominal cash flows do not pollute our

estimation of stock price informativeness.

1.2 Measuring Price Informativeness

While there is a debate in the empirical literature about how to best measure price informativeness (e.g. Philippon, 2015), the measure suggested by Bai, Philippon and Savov (2016) is closest to our model's measure. It captures the extent to which asset prices in year t are able to predict future cash-flows in year $t+k$.

Their informativeness measure is constructed by running cross-sectional regressions of future earnings on current market prices. Controlling for other observables limits the risk of confounding public information impounded in prices with markets foresight. For each firm j, in year t, we estimate k-period ahead informativeness as

$$\frac{E_{j,t+k}}{A_{j,t}} = \alpha + \beta_t log\left(\frac{M_{j,t}}{A_{j,t}}\right) + \gamma X_{j,t} + \epsilon_{i,t}, \tag{1}$$

where $E_{j,t+k}/A_{j,t}$ is the cash-flow of firm j in year $t+k$, scaled by total assets of the firm in year t; $\log(M_{j,t}/A_{j,t})$ is firm market capitalization, scaled by total assets; and $X_{j,t}$ are controls for firm j that capture publicly available information. In the main specification, the controls are current earnings and industry sector (SIC 1) fixed effects. When we estimate price informativeness at the industry level (SIC3 or SIC2), we need to drop the industry fixed effect as a control.

The parameter $\beta_{t,k}$ measures the extent to which firm market capitalization in year t can forecast the firm cash-flow in year $t+k$. To map this coefficient into a proxy of price informativeness, we follow Bai et al. (2016) and do the following adjustment:

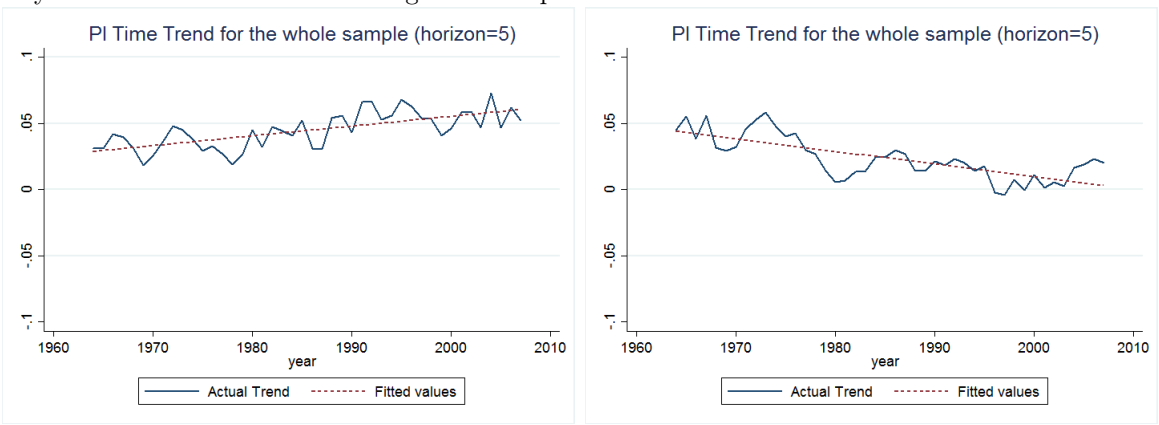
$$\left(\sqrt{P_{Info}}\right)_t = \beta_t \times \sigma_t\big(log(M/A)\big) \tag{2}$$

where $\sigma_t\big(log(M/A)\big)$ is the cross-sectional standard deviation of the forecasting variable log(M/A) in year t. The use of square root ensures that we can directly interpret economically the measure as it gives us the dollars of future cash flows per dollar of current total assets.

1.3 Aggregate Trends in Price Informativeness

We first establish the empirical puzzle that motivates our analysis. Price informativeness increases over time for firms in the S&P 500 (Bai, Philippon, and Savov (2013)'s headline result), but it decreases for the whole sample of publicly-listed firm. Figure 1 illustrates the contrast between the increase in informativeness for S&P 500 firms (left panel) and the decrease in price informativeness for all firms (right panel).

Figure 1: **Price Informativeness is Rising for S&P 500 Firms and Falling for All Public Firms**. Results from the cross-sectional forecasting regression (eqn 1): $E_{i,t+k}/A_{i,t} = \alpha + \beta_{t,k} log(M_{i,t}/A_{i,t}) + \gamma X_{i,t} + \epsilon_{i,t,k}$, where M is market cap, A is total asset, E is earnings before interest and taxes($EBIT$) and X are a set of controls that captures information publicly available. We run a separate regression for each year $t = 1964, ..., 2008$ and horizon $k = 5$. Price informativeness is $\beta_{t,k} \times \sigma_t(log(M/A))$, where $\sigma_t(log(M/A)$ is the cross-sectional standard deviation of $log(M/A)$ in year t. Above each plot is a linear trend normalized to zero and one at the beginning and end of the sample (plotted in dashed lines). The left figure contains S&P 500 nonfinancial firms from 1964 to 2008, while the right figure contains all publicly listed nonfinancial firms during the same period.



Similar plots in the Appendix reveal that the trends are nearly identical for 3-year and 5-year horizons. Therefore we proceed looking only at 5-year price informativeness.

Table 1 quantifies these trends and demonstrates the statistical significance of the difference between the S&P 500 and all-public-firm samples. Both trends are economically large. For the S&P 500 sample, the mean of price informativeness is 4.1 and its time-series standard deviation is 1.1. Between 1964 and 2008, price informativeness rose 70% relative to its mean, or 2.1 standard deviations. For non-S&P 500 firms, the average level of price informativeness is 2.8 with a time-series standard deviation of 1.2. So the fall in price informativeness is 100% relative to the mean and 2.5 times the standard deviation.[1]

[1]For the S&P 500 sample, the interquartile range in price informativeness is 1.62. The rise in price

Table 1: Price Informativeness Trends over Time

Dep. Var	100× Price Informativeness			
Sample	S&P 500		All Listed Firms	
Horizon	h=3	h=5	h=3	h=5
	(1)	(2)	(3)	(4)
Time Trend	1.76***	2.38***	-3.11***	-2.96***
	(0.33)	(0.41)	(0.49)	(0.52)
Observations	45	45	45	45

This table presents time series regressions of price informativeness by horizon. Price informativeness is calculated as in Eq. 2 using estimates from the cross-sectional forecasting regression 1. For this table, we regress the time series of price informativeness at a given horizon $h = 3$, 5 years on a linear time trend normalized to zero and one at the beginning and end of the sample. Newey - West standard errors, with five lags are in parentheses. *** denotes significance at the 1% level.

2 Where Is Information Flowing?

The divergent aggregate informativeness trends offer a puzzling and mixed message about whether the financial sector is becoming more efficient or not. To understand what is going on and why, this section cuts the sample of firms in different ways, to understand which prices are becoming more informative and which less, or if this is a composition effect.

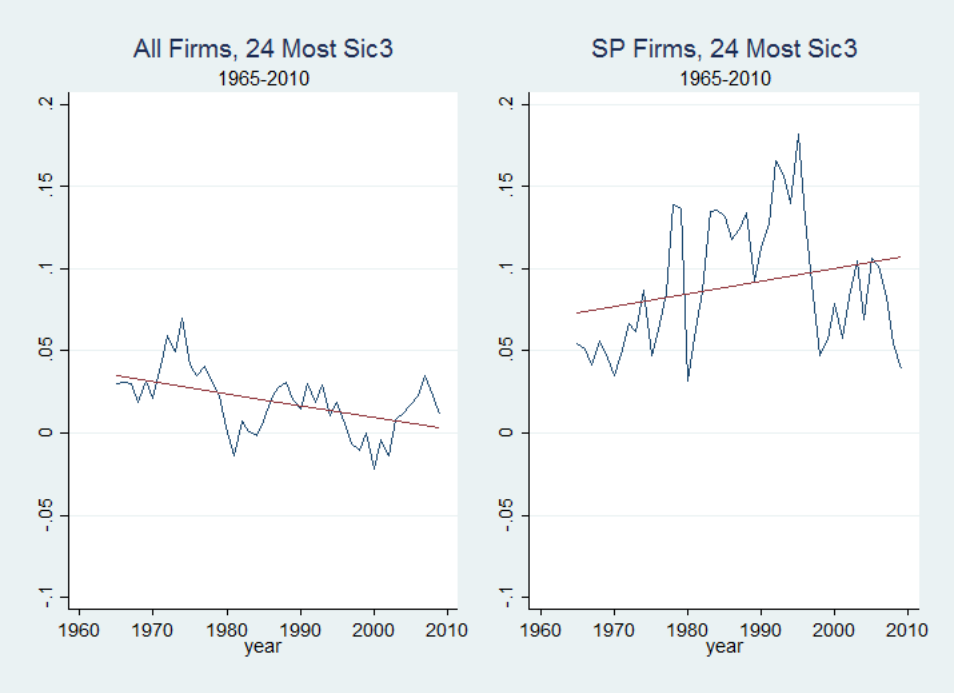
Is information flowing to specific industries? One plausible explanation is that the market is getting better at pricing some types of firms. Perhaps health care or online firms were hard to price initially, or this is a story about reasearch and development intensity of firms. These features are all highly correlated with a firm's industry. So, we begin by asking if the growth or decline in price informativeness is determined by an industry effect.

There are 253 different SIC3 codes in Compustat and 173 in S&P 500. The median number of firms per industry is 12, but the distribution is very skewed. Looking at the industries with strictly more than the median number of firms, we end up with only 24 distinct industries. We call these 24 industries *SPindustries*. Then, we restrict our sample

informativeness is about two times this interquartile distance. For non-S&P 500 firms, the interquartile range is 1.1. The fall in price informativeness is more than twice this interquartile distance.

only to firms in these 24 SPindustries. Within this restricted sample, we compare price informativeness trends of firms that appear in the S&P 500, at some point (542 firms) with those in the same industries, that do not (7,768 firms).

Figure 2: **Price Informativeness for S&P 500 Industries is Stagnant**. Left panel shows non-S&P 500 firms in the 24 "SPindustries", defined as industries most represented in the S&P. The right panel shows S&P 500 firms in the same set of industries. Price informativeness is obtained separately for each group by running the forecasting regression (eqn 1) for horizon $k = 5$ and then using Eq 2.



The left side of Figure 2 shows that non-S&P 500 firms in S&P 500 industries do not experience a rise in price informativeness. The right side shows that S&P 500 firms in these same industries do experience the improvement in price efficiency.

If we do the same exercise with every industry represented in the S&P, instead of just the 24 most represented industries, we get similar results. This difference in price informativeness does not appear to be driven by differences in industries. This evidence suggests that the increase in price information for S&P 500 firms does not result from S&P 500 firms being in more informative industries.

Is information flowing specifically to firms currently in the S&P 500? No, this does not seem to be a result about firms currently in the S&P 500 having greater price

informativeness or a different trend. Instead, the rise in price informativeness seems to affect the type of firm that would be in the S&P 500. We show this result in two stages and then continue to investigate the question of what firm characteristics determine rising or declining price informativeness.

To look at the question of whether there is something specific to firms in the S&P 500, we perform two different tests. First, we look at firms, which at some point will be part of the S&P 500, and compare their price informativeness trend, (a) during the period where they are in the S&P 500; and (b) during the period where they are not. Second, we look at the price informativeness evolution of firms that are similar to firms in the S&P 500.

For the first exercise, we estimate two separate estimations of Equation 1 for the period of the firm life when it is in the S&P 500 and when it is not. Figure 3 shows that, among the sample of firms that are in the S&P 500 at some point in their life, the trend in price informativeness is similar for firms currently in and out of the S&P 500 . In levels, informativeness is higher when a firm is not in the S&P 500, than when they are in.

Figure 3: **Price Informativeness Trend While in and out of S&P 500 is Similar**. The sample for both lines contains publicly listed nonfinancial firms that have been in the S&P 500 at some time between 1964 and 2008. 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 currency in the S&P 500 . The black and red dashed lines are linear trends that fit the grey and red time trends, respectively. Price informativeness is obtained separately for each group by running the forecasting regression (eqn 1) for horizon $k = 5$ and calculating the product of the forecasting coefficient and the cross-sectional standard deviation of market prices in year t using eqn 2.

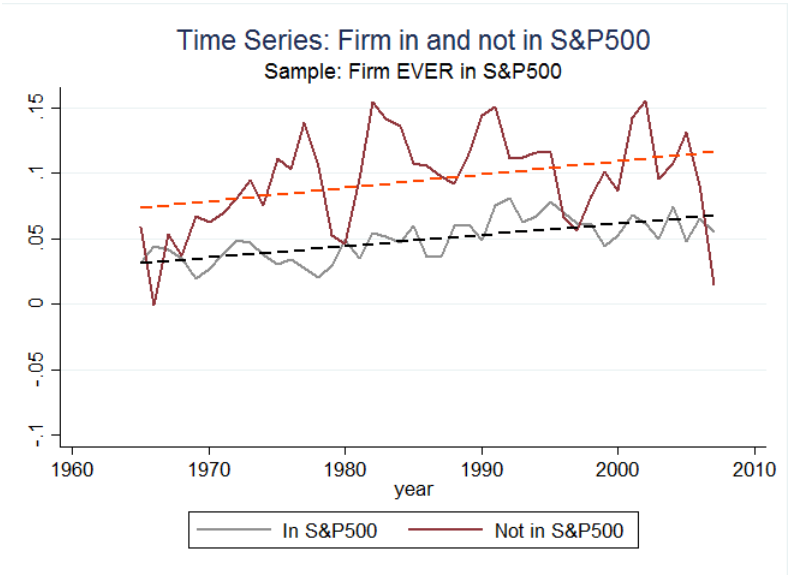
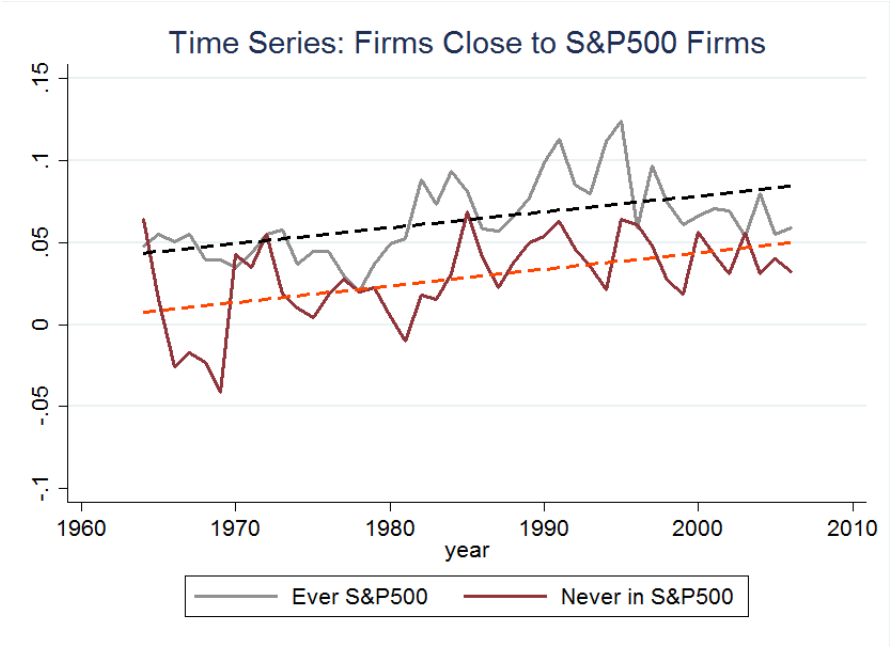


Figure 4: **Price Informativeness Grows Similarly for Firms Similar to S&P 500**. Price informativeness is the ability to forecast future earnings (eqn 2). The future earnings are measured here at 5-year horizon. The sample is all firms whose probability of being in the S&P 500 , as determined by (3), exceeds 0.6.



Next, we compute informativeness for each subsample using Equation 1. Figure 4 shows that firms that will never be in the S&P 500 but are relatively close in term of market capitalization and size exhibit a smaller level of informativeness and a slightly smaller rise over time. While the difference in price informativeness is not statistically significant at the start of the sample, by the mid-1980's the confidence intervals do not overlap. The difference between the S&P 500 and non-S&P 500 firms at the end of the sample is a significant difference.

To investigate whether firms with similar characteristics have similar changes in stock price informativeness, we proceed in two steps. First, for the whole sample of firms at each date, we construct a dummy variable $\tilde{SP500}_{i,t}$, which takes the value of one if firm i is in the S&P 500 at time t and zero otherwise. Then, we estimate α, δ, ϕ and γ in the the following equation:

$$\tilde{SP500}_{i,t} = \alpha_i + \delta_t + \phi_t log(M/A) + \gamma_t log(Asset) + \epsilon_{i,t} \tag{3}$$

Next, we use the estimates of α, δ, ϕ and γ to construct predicted probabilities of being

in the S&P 500. We denote this probability as $\widehat{SP500}_{i,t}$. The median score of $\widehat{SP500}_{i,t}$ for firms in the S&P 500 is around 0.6. All firms with $\widehat{SP500}_{i,t} \geq 0.6$ we call *similar* to S&P 500 firms. For the next exercise, we restrict the sample to on similar firms. This leaves 3,105 distinct firms, among which 60% will be indeed at some in the S&P 500 and 40% will not.

We learn that there is something about the type of firm in the S&P 500, the size or book-to-market, that draws in more analysis over time. For some reason, firms with similar characteristics that will never be S&P 500 firms have lower informativeness levels, but a similar informativeness growth rate.

Are older / bigger firms attracting all the data? Another possible explanation is that firms in the S&P 500 are, on average, much older / bigger than other firms. Could differences in firm age & size explain the different trends in informativeness? Perhaps big data enabled us to improve analysis of old firms more than new ones? This hypothesis holds more promise. There are systematic differences in the level and trend of informativeness between young and old firms. But, this does not explain all of the difference between S&P 500 and non-S&P 500 firms.

We compute price informativeness into seven age bins in the following way: For the whole sample period, we compute bins based on firm age (defined as the first time a firm is listed in CRSP). Bins are defined such that they contain roughly the same number of observations to avoid having biased estimates coming from large differences in sample size.

Then we run separate cross-sectional regressions of the following form:

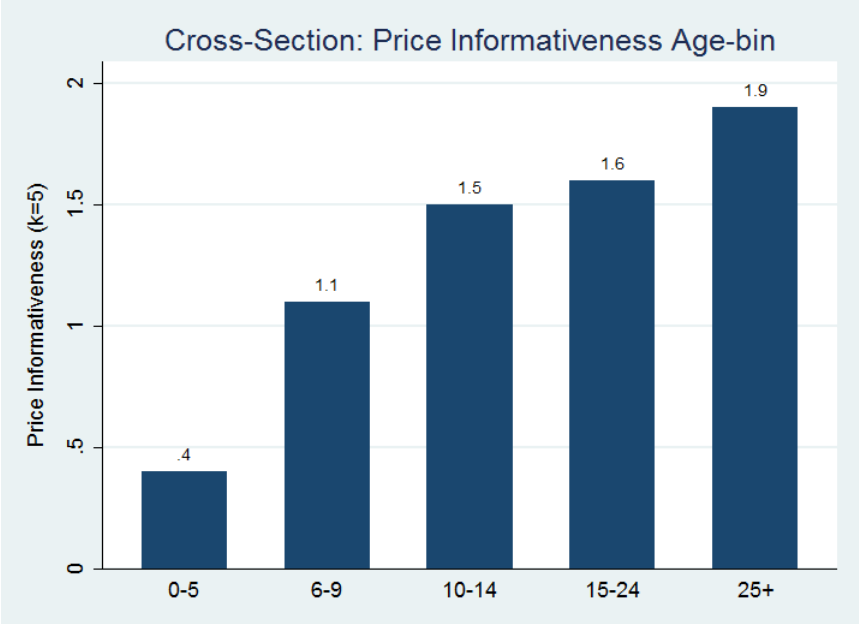
$$\frac{E_{i,y,t+k}}{A_{i,y,t}} = \alpha + \beta_{t,y} log\left(\frac{M_{i,y,t}}{A_{i,y,t}}\right) + \gamma X_{i,y,t} + \epsilon_{i,y,t}$$

where $E_{i,y,t+k}/A_{i,y,t}$ is the cash-flow of firm i belonging to age bin y in year $t+k$ scaled by total asset of the firm in year t.[2]

Figure 5 shows that older firms have more informative prices. This effect is more pronounced for firms that move from early to middle age. For the oldest firms, age matters less.

[2]Adding Year fixed effects to the cross-sectional specification does not change the result.

Figure 5: **Price Informativeness for Old Firms is Higher**. Price informativeness is the ability to forecast future earnings (Eq 2). We run a separate regression for each year $t = 1964, ..., 2008$, horizon $k = 5$ and age interval $[1/5), ...[5/5]$ partitioned by $1/5$ age quantiles. Price informativeness is the average value of $\beta_{t,y} \times \sigma_{y,t} log(M/A)$, where $\sigma_{y,t} log(M/A)$ is the cross-sectional standard deviation of $log(M/A)$ in year t and age interval y. Future earnings are measure here at 5-year horizons. The sample contains publicly listed nonfinancial firms from 1964 to 2008.



Nearly identical results emerge when we split firms based on firm size (see Appendix). Size is defined as firm total assets, deflated. We then split firms into seven size bins. In the cross-section, stock price informativeness increases with size for the first four bins and then plateaus.

Taken together, these results teach us that the increase in price information for S&P 500 firms may arise because of a change in composition of the S&P 500 . Given that old / large firms have more informative prices, a change in the composition of the S&P 500 toward larger / older firms can explain why the S&P 500 is becoming more informative, even if the informativeness of the oldest / largest firms is not rising. We explore this possible composition effect next.

Is this a composition effect? The next possibility we investigate is that perhaps financial markets are not getting better at pricing older (bigger) firms over time, or any kind of firm in particular. It's simply that young / small firms have always been hard to price accurately

and the composition of the S&P 500 changed so that there are fewer young / small firms in the index. In other words, S&P 500 price efficiency is rising because the average S&P 500 firm is getting older / bigger. For this composition effect to explain the decline in overall price efficiency for all firms, it would have to be that the average non-S&P 500 firm is getting younger / smaller.

Figure 6: **S&P 500 firms become older and bigger. Non-S&P 500 firms do not** We compare the average age and size of firms that were ever in S&P 500 and firms there were never in S&P 500 over time. Firm size is defined as firm's total asset. The sample contains publicly listed nonfinancial firms from 1960 to 2010.

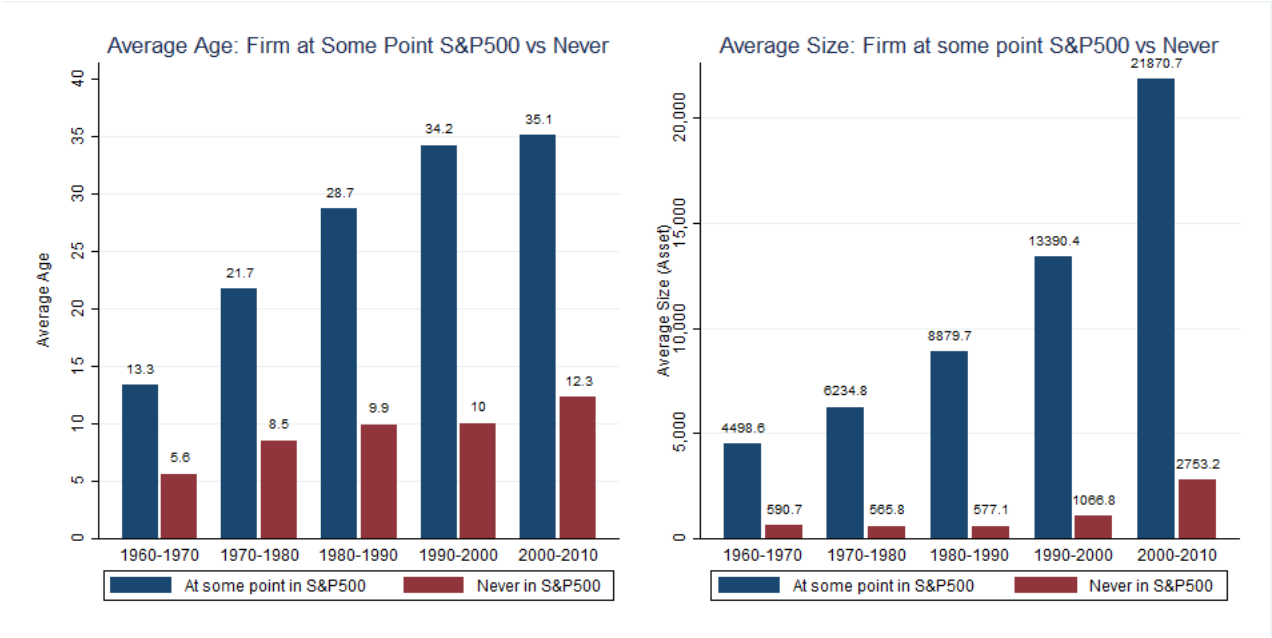
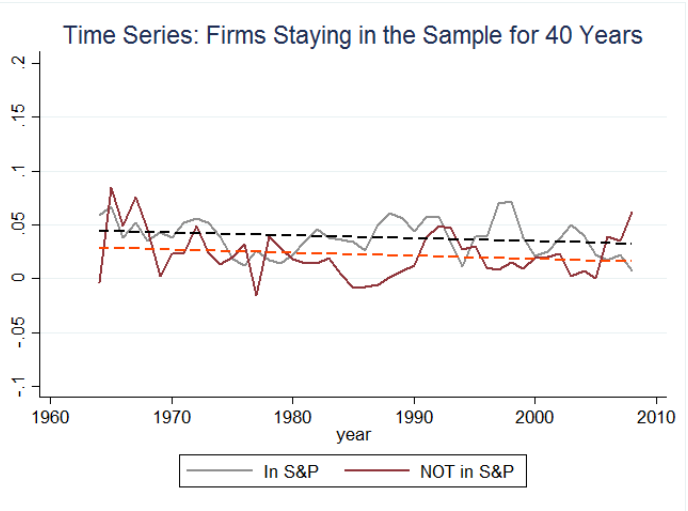


Figure 6 supports the first hypothesis that S&P 500 firms are getting older and larger. But it does not support the second hypothesis that non-S&P 500 firms are getting younger. When we do the same breakdown by size (right figure), we see again that S&P 500 firms are growing much larger, relative to non-S&P 500 firms.

If we remove the age composition effect, by only considering long-lived firms, the trends in price informativeness go away, for both S&P 500 and non-S&P 500 firms.

The firms in Figure 7 remain listed for at least 40 years, in the sample. That means they are all live for almost the whole sample period. So, this result is mostly composition-bias free. Without a change in the sample of firms, price informativeness is constant for both

Figure 7: **Without Composition Effect, Price Informativeness is Stagnant**. Price informativeness is the ability to forecast future earnings (Eq 2) at a 5-year horizon. We compare price informativeness for firms that were at some point in S&P 500 and existed for at least 40 years (grey line) and firms that were never in S&P 500 and existed for at least 40 years (red line). The black and red dashed lines are linear trends that fit the grey and red time trends, respectively. Price informativeness is obtained separately for each group. The sample contains publicly listed nonfinancial firms that existed for at least 40 years from 1964 to 2008.

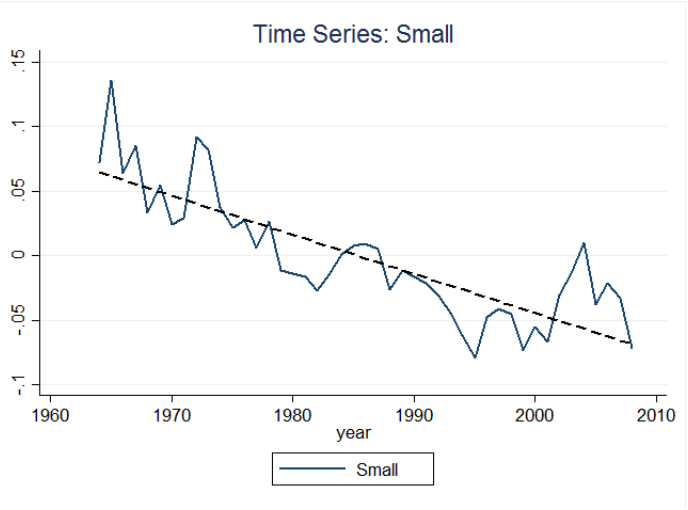


categories of firms. Of course, these firms are all aging over the course of the sample. So this is not holding the current age of the firm fixed. But since older firms should have more informative prices, that makes the lack of a trend here all the more striking. This result lends more support to the idea that the increase in price informativeness for S&P 500 firms is mostly the result of a change in composition toward older, larger firms, rather than an increase in the informativeness of the oldest / largest firms.

Price informativeness of newest, smallest firms is declining. While the change in composition of the S&P 500 firms is much of the story, it's not the only trend driving price informativeness of old firms and young firms apart. Recall that price informativeness of the whole sample was falling. That effect comes not from composition changes, but from a decline in the price informativeness of new entrants and the smallest, youngest firms.

Figure 8 shows that the decline in price informativeness comes exclusively from small firms. To estimate the result, we construct two bins of equal numbers of firms, grouped by size. We split firms between small and big where small means a firm is below the median of the distribution at that point in time. Using stricter definition of small, such as below the

Figure 8: **Price Informativeness of Small Firms is Falling**. For each year, the sample is made of all publicly listed nonfinancial firms below the median size at that year, as measured by total asset. The sample contains all publicly listed nonfinancial firms, which assets are below the median at that year. Price informativeness is the ability to forecast future earnings (Eq 2) at a 5-year horizon and estimating by running a separate regression for each year $t = 1964, ..., 2008$.



top quartile yields similar results. We then estimate Equation 1 on small and large firms separately. We plot time series evolution of price informativeness for small and big firms. The trend over time is clearly negative for young firms. However, for the larger firms, price informativeness is flat and perhaps slightly increasing over time.

If we do a similar sort by firm age, we get similar decline for the youngest firms. Specifically, the age results splits the sample between firms less than versus more than 10 years old. We then estimate Equation 1 on each subsample separately to see the time series evolution of informativeness for young and old firms.

It is possible that even among the youngest firms, there is a composition effect coming from differences in cohort.[3] When we estimate the price informativeness of newly entering firms only, we also find a decline in their price informativeness. The result is based on estimates of 5 cohorts, each spanning 10 years. Then we estimate price informativeness in the year of firm entry, for all the firms that enter within the 10-year cohort window. We do this with and without year fixed effects. In both cases, the price informativeness of newly entering firms declines over time.

[3]The idea being that a 1-year old firm in 1970 may be different than a 1-year old firm in 1980.

Our conclusion is that while the growth in price informativeness among S&P 500 firms is an age or firm size composition effect, the decline in price informativeness for the smallest firms is not. For some reason, the average investor seems to have less precise information about the future earnings of small firms today, than they did in decades past.

3 A Model to Interpret Patterns In Price Information

The data reveal two opposing trends: an increase in the informativeness of S&P 500 firms, that appears to be driven by a composition effect, and a decline in the informativeness of small firms. Given the growth in computing speed, data availability, and human resources devoted to the financial sector, it is puzzling that many firms are priced less accurately today than such firms were in the past. It also leaves open the question: Is this efficiency-enhancing? After all, the largest firms, that constitute most of the market, are priced more efficiently. Or is there a welfare loss from information processing resources devoted to a small number of already well-understood firms? Without a model, data alone cannot tell us if another allocation might be better and why.

We use a model to explore what long-run trends might possibly explain our documented facts: Over time, investors want to process more data about large firms and less data abut small ones. We learn that the growth in the size of large firms is essential to explain the decline of small firm data processing. It is possible that the long-run productivity of the financial sector has also improved, but only modestly. We then ask the model whether this trend is costly for social welfare. We learn why information externalities lead investors to naturally process more big-firm data than is socially optimal.

To deliver these insights, we need an equilibrium model with multiple assets and agents who choose how much data to process about each asset. We start with a framework like that in Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016). If we assumed, exogenously, that information processing is directed at particular assets, it would not explain why some prices are becoming more efficient and others are not. Instead, this framework allows the model to predict where the information should flow. The model teaches us about how a profit maximizing investor should use data processing and invest, and how this should affect

the information contained in equilibrium prices.

To this framework, we add real spillovers that can speak to social efficiency. Our stylized toy model of the real economy is designed to show one possible reason why financial price informativeness might have economic consequences. In this case, commonly-used compensation contracts that tie wages to firm equity prices (e.g., options packages) also tie price informativeness to optimal effort. Since investors are infinitesimal and take prices as given, they do not internalize the effect of their information and portfolio choices on manager's decision through price informativeness. The fact that there are economic externalities is by construction. The result that the social planner favors more data processing about small firms is not.

Firm Manager's Problem This is a repeated, static model in discrete time. We use t to denote time. But the only thing changing over time is information technology or the size of the largest firm. At each date t, there is a one-period problem for each firm manager and each investor to solve.

The key friction is that the manager's effort choice is unobserved by investors. The manager exerts costly effort only because he is compensated with equity, whose value is responsive to his effort. Because asset price informativeness governs the responsiveness of price to effort, it also determines the efficiency of the manager's effort choice.[4]

The profit of each firm j, $\tilde{d}_{jt}$, depends on the firm manager's effort, which we call labor l_{jt}. Specifically, the payoff of each share of the firm is $\tilde{d}_{jt} = g(l_{jt}) + \tilde{y}_{jt}$, where $g(l) = l^{\phi}$, $\phi \leq 1$ is increasing and concave and the noise $\tilde{y}_{jt} \sim N(0, \tau_0^{-1})$ is i.i.d. and unknown at t. Because effort is unobserved, the manager's pay w_{jt} is tied to the equity price p_{jt} of the firm: $w_{jt} = \bar{w}_j + p_{jt}$. However, effort is costly. We normalize the units of effort so that a unit of effort corresponds to a unit of utility cost. Insider trading laws prevent the manager from participating in the equity market. Thus, each period, the manager chooses l_{jt} to maximize

$$U_m(l_{jt}) = \bar{w}_j + p_{jt} - l_{jt} \tag{4}$$

[4]Of course, this friction reflects the fact that the wage is not an unconstrained optimal contract. The optimal compensation for the manager is to pay him for effort directly or make him hold all equity in the firm. We do not model the reasons why this contract is not feasible because it would distract from our main point.

Each period, the firm j pays out all its profits $\tilde{d}_{jt}$ as dividends to its shareholders. We let $\tilde{d}_t$ denote the vector whose j^{th} entry is $\tilde{d}_{jt}$

Assets The model features 1 riskless and n risky assets. The price of the riskless asset is normalized to 1 and it pays off r at the end of each period. One share of a risky asset is a claim to the random payout $\tilde{d}_{jt}$ at the end of the period. For simplicity, we assume that these asset payoffs are independent. The riskless asset pays a known amount $1 + r$ at the end of the period.

There are n risky assets, one for each of the firms in the economy. Each share of a risky asset j is a claim to the payoff $\tilde{d}_{jt}$. Each risky asset has a stochastic supply given by $\bar{x}_j + x_{jt}$, where noise x_{jt} is normally distributed, with mean zero, variance σ_x, and no correlation with other noises: the vector of x_{jt}'s is $x_t \sim \mathcal{N}(0, \sigma_x I)$. As in most noisy rational expectations equilibrium model, the supply is random to prevent the price from fully revealing the information of informed investors. This randomness might be interpreted as investors in the market trading for hedging reasons that are unrelated to information, as in Manzano and Vives (2010).

Portfolio Choice Problem Each period, a new continuum of atomless investors is born. Each investor is endowed with initial wealth, W_0.[5] They have mean-variance preferences over ex-post wealth, with a risk-aversion coefficient ρ. Let E_i and V_i denote investor i's expectations and variances conditioned on all interim information, which includes prices and signals. Thus, investor i chooses how many shares of each asset to hold, q_{it} to maximize interim expected utility, $\hat{U}_{it}$:

$$\hat{U}_{it} = \rho E[W_{it}|\mathcal{I}_{it}] - \frac{\rho^2}{2} V[W_{it}|\mathcal{I}_{it}] \tag{5}$$

subject to the budget constraint:

$$W_{it} = rW_0 + q'_{it}(\tilde{d}_t - p_t r), \tag{6}$$

where q_{it} and p_t are $n \times 1$ vectors of prices and quantities of each asset held by investor i.

[5]Since there are no wealth effects in the preferences, the assumption of identical initial wealth is without loss of generality. The only consequential part of the assumption is that initial wealth is known.

Data Processing Choice Investors can acquire information about asset payoffs $\tilde{d}_t$ by processing digital data. Digital data is coded in binary code. Investors face a constraint $\bar{B}_t$ on the total length of the binary code they can process. This constraint represents the frontier information technology in period t. One $\{0, 1\}$ digit encodes 1 bit of information.[6] Thus units of binary code length are bits.

All data processing is subject to error. The most common model of processing error is the parallel Gaussian channel.[7] For a Gaussian channel, the number of bits required to transmit a message is related to the signal-to-noise ratio of the channel. Clearer signals can be transmitted through the channel, but they require more bits. The relationship between bits and signal precision for a Gaussian channel is $bits = 1/2log(1 + \text{signal-to-noise})$ (Cover and Thomas (1991), theorem 10.1.1).

Investors choose how to allocate their capacity among n risky assets. Let b_{it} be a vector whose jth entry, $b_{it}(j) > 0$, is the number of bits processed by agent i at time t about $\tilde{d}_{jt}$. Let η^b_{it} represent the realized string of zeros and ones that investor i observes. The data processing constraint is then

$$\sum_{j=1}^{N} b_{it}(j) \leq \bar{B}_t \qquad \text{where} \quad b_{it}(j) \geq 0 \quad \forall i,\, j,\, t. \tag{7}$$

Information sets and equilibrium The information set the investor has when he makes investment decisions is $\mathcal{I}_t = \{\mathcal{I}_{t-1}, \eta^b_{it}, p_t\}$. The ex-ante information set includes the entire sequence of data processing capacity: $\mathcal{I}_0 \supset \{\bar{B}_t\}_{t=0}^{\infty}$.

An equilibrium is a sequence of effort choices by managers $\{l_t\}$, precision choices, $\{K_{it}\}$, which are diagonal and positive semi-definite, and portfolio choices $\{q_{it}\}$ by investors such that

1. Firm managers' effort decision maximizes (4), at each date t.

[6] A byte is 8 bits, which allows for 256 possible sequences of zeros and ones, enough for one byte to describe an alpha-numeric character or common keyboard symbol. Megabytes are 10^6 bytes. If your computer can store 1GB in its RAM, that is 10^9 bytes, or a binary code of length 8×10^9.

[7] As Cover and Thomas (1991) explain, "The additive noise in such channels may be due to a variety of causes. However, by the central limit theorem, the cumulative effect of a large number of small random effects will be approximately normal, so the Gaussian assumption is valid in a large number of situations."

2. Investors choose bit string lengths $b_{it} \geq 0$ to maximize $E[\hat{U}_{it}|\mathcal{I}_{t-1}^{+}]$, where $\hat{U}_{it}$ is defined in (5), taking the choices of other agents as given. This choice is subject to (7).

3. Investors choose their risky asset investment q_{it} to maximize $E[U(c_{it})|\eta_{fit}, p_t]$, taking asset prices and others' actions as given, subject to the budget constraint (6).

4. At each date t, the vector of equilibrium prices p_t equates aggregate demand (left side) with supply (right) to clear the market:

$$\int_i q_{ijt} di = \bar{x}_{jt} + x_{jt}, \tag{8}$$

3.1 Solving the Model

We solve the model in five steps. We sketch each step here and relegate details to the appendix for the interested reader. Because units of signal precision are easier to work with than bits, we define K_{ijt} to be the precision of the signal η_{ijt} inferred from the data processed by investor i about firm j at time t. Let K_{it} be the diagonal matrix with K_{ijt} on its jth diagonal and η_{it} be the vector of all signals observed by i. Finally, let $\bar{K}_t \equiv \int_i K_{it} di$ be the matrix of the average investors' signal precision.

Step 1: Bayesian updating. There are three types of information that are aggregated in time-2 posteriors beliefs: prior beliefs, price information, and (private) signals from data processing. We begin with price information. We conjecture and later verify that a transformation of prices p_t generates an unbiased signal about the risk factor payoffs, $\eta_{pt} = \tilde{d}_t + \epsilon_{pt}$, where $\epsilon_p \sim N(0, \Sigma_p)$, for some diagonal variance matrix Σ_p.

Next, we construct a single signal that encapsulates the information conveyed in bit strings. Recall that in a Gaussian channel with prior information precision τ, the number of bits required to transmit a signal with a given precision Ω is $bits = 1/2 \cdot log(1 + \tau_0^{-1} K_{it})$. The data contains the true value of $\tilde{y}_t$. But data processing is imperfect and introduces Gaussian noise. Processed fundamental data is observed as $\eta_{fit} = \tilde{y}_t + \tilde{\epsilon}_{fit}$, where the channel (data processing) noise is a normal, random variable: $\tilde{\epsilon}_{fit} \sim N(0, K_{it}^{-1})$. Substituting this mapping into (7) yields a new data processing constraint in terms of signal precisions $K_{it} \geq 0$:

$$\sum_{j=1}^{N} log(1 + \tau_0^{-1} K_{it}) \leq 2\bar{B}_t. \tag{9}$$

Finally, Bayes' law tells us how to combine price signals, data signals and prior beliefs. The resulting posterior beliefs about z are normally distributed with variance $var[\tilde{d}_t|\mathcal{I}_{it}] = (\Sigma^{-1} + \Sigma_p^{-1} + K_{it})^{-1}$ and mean

$$E[\tilde{d}_t|\mathcal{I}_{it}] = var[\tilde{d}_t|\mathcal{I}_{it}](\Sigma^{-1}\mu + K_{it}\eta_{it} + \Sigma_p^{-1}\eta_{pt}). \tag{10}$$

Step 2: Solve for the optimal portfolios, given information sets and issuance. Substituting the budget constraint (6) into the objective function (5) and taking the first-order condition with respect to q_{it} reveals that optimal holdings are increasing in the investor's risk tolerance, precision of beliefs, and expected return:

$$q_{it} = \frac{1}{\rho} var[\tilde{d}_t|\mathcal{I}_{it}]^{-1}(E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r). \tag{11}$$

Step 3: Clear the asset market. Substitute each agent j's optimal portfolio (11) into the market-clearing condition (8). Collecting terms and simplifying reveals that the vector of equilibrium asset prices are linear in payoff risk shocks and in supply shocks:

$$p_t = A_t + C_t(g(l_t) + \tilde{y}_t) + D_t\tilde{x}_t \tag{12}$$

where $\tilde{y}_t$ is the vector of unexpected shocks to dividends and $\tilde{x}_t$ is the vector of asset supply shocks at time t. Coefficients A, C, and D are in the Appendix. The vector of all managers' effort shows up as in price as $+C_t g(l_t)$. Since the manager gets paid with equity, whose value is p_t, the size of C_t governs the manager's incentive to exert effort.

Step 4: Solve for data processing choices. The information choice objective comes from substuting in the optimal portfolio choice and equilibrium price rule, and then taking the ex-ante expectation over the signals and price that are not yet observed at the start of the period. This yields an objective that is linear in signal precisions;

$$\max_{K_{i1t},\ldots,K_{int}\geq 0} \sum_{j=1}^{n} \Lambda(\bar{K}_{jt}, \bar{x}_j) K_{ijt} + \text{constant} \tag{13}$$

$$\text{where} \quad \Lambda(\bar{K}_{jt}, \bar{x}_j) = \bar{\sigma}_j[1 + (\rho^2 1/\tau_x + \bar{K}_{jt})\bar{\sigma}_i] + \rho^2 \bar{x}_j^2 \bar{\sigma}_j^2, \tag{14}$$

and $\bar{\sigma}_j^{-1} = \int \hat{\Sigma}_i^{-1}(j,j)di$ is the average precision of posterior beliefs about asset j. Its inverse, average variance $\bar{\sigma}_j$ is decreasing in $\bar{K}_{jt}$. The appendix shows two important properties. The first is strategic substitutability in data choices: $\partial\Lambda(\bar{K}_{jt}, \bar{x}_j)/\partial\bar{K}_{jt} < 0$. The second is returns to asset scale in data processing: $\partial\Lambda(\bar{K}_{jt}, \bar{x}_j)/\partial\bar{x}_j > 0$.

Maximizing a weighted sum (13) subject to a concave constraint (9) yields a corner solution. The investor optimally processes data about only one asset. Which asset to learn about depends on which has the highest marginal utility $\Lambda(\bar{K}_{jt}, \bar{x}_j)$. If there is a unique asset $j^* = \text{argmax}_j \Lambda(\bar{K}_{jt}, \bar{x}_j)_j$, then the solution is to set $K_{i,j^*,t} = \tau_0(e^{\bar{B}_t} - 1)$ and $K_{ilt} = 0$, $\forall l \neq j^*$. But when capacity $\bar{B}_t$ is high enough, there will exist more than one asset j that is learned about. Let $\mathcal{M}_t \equiv \left\{\bar{K_{jt}} > 0\right\}_{j=1}^{n}$ be the set of assets learned about. Then an equilibrium is a set of average precisions for each asset j, $\left\{\bar{K_{jt}}\right\}_{j=1}^{n}$ such that

$$\Lambda(\bar{K}_{jt}, \bar{x}_j) = \bar{\Lambda} \qquad \forall j \in \mathcal{M}_t \tag{15}$$

In this equilibrium, investors are indifferent about which single asset $j \in \mathcal{M}$ to learn about. But the aggregate allocation of data processing is unique (Kacperczyk, Van Nieuwerburgh, and Veldkamp, 2016).

Step 5: Solve the firm manager's problem. The firm manager chooses his effort to maximize (4). The first order condition is $C_{jt}g'(l_{jt}) = 1$, which yields an equilibrium effort level $l_{jt} = (g')^{-1}(1/C_{jt})$ for $j = 1, 2$.

Recall that $g(l) = l^\phi$, $\phi \leq 1$ is the increasing, concave production function of firm manager. Using the manager effort first order condition $g'(l) = \frac{1}{C}$, we find that optimal effort is

$$l = (\phi C)^{\frac{1}{1-\phi}} \tag{16}$$

The link between data choices and real efficiency. A reason for writing down this model is that it can teach us about the social efficiency of investors' data processing choices. The key link between data processing and social welfare is the relationship of each with price informativeness.

From the asset market solution, we work out a function Γ that maps data processing precision $\bar{K}_{jt}$ into assets' price informativeness. To do this, we start with the equilibrium price solution, $\frac{C_{jt}}{D_{jt}} = \frac{\bar{K}_{jt}}{\rho}$. We then substitute that price coefficient ratio in the posterior variance expression to get $var[g(l_{jt}) + \tilde{y}_{jt}|\mathcal{I}_{ijt}] = \tau_{j0} + \bar{K}_{jt} + (C_{jt}/D_{jt})^2 \tau_{xj} = \tau_{0j} + \bar{K}_{jt} + \left(\bar{K}_j/\rho\right)^2 \tau_{xj}$. Then, we substitute this conditional variance in the formula (27) for C to get mapping from data processing to price informativeness.

$$C_{jt} \equiv \Gamma(\bar{K}_{jt}; \Theta) = 1 - \frac{1}{1 + \frac{\bar{K}_{jt}}{\tau_{0j}} + \frac{\tau_{xj}}{\tau_{0j}} \left(\frac{\bar{K}_{jt}}{\rho}\right)^2} \tag{17}$$

where $\Theta = \{\rho, \tau_{0j}, \tau_{xj}\}$ is the vector of underlying parameters.

Thus, as more information is analyzed, C_{jt} rises and managers are better incentivized to exert optimal effort. While the model is stylized, it is designed to illustrate one of many possible reasons why trends in financial analysis matter for the real economy.

3.2 Understanding Trends in Price Informativeness

We start by setting up the puzzle that motivates the paper. If faster computers can process ever more data over time, why haven't all firms prices benefited from the increase in price informativeness? We show that, although investors prefer to learn about large firms, more data does not make them want to learn less about small firms. Instead, all firms should experience an increase in price informativeness. Thus an increase in the efficiency of the financial sector in processing data is not a complete explanation for the trends in the data.

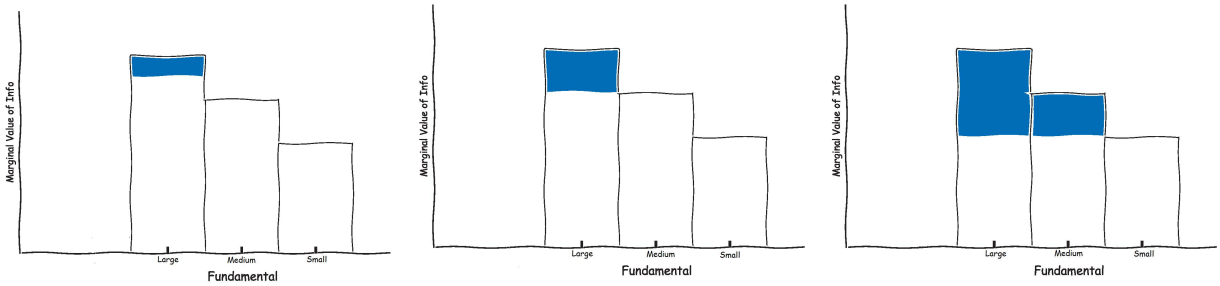
The second set of results offers a solution. It shows that if large firms grow larger, as they do in our data, this trend alone can explain the composition effect driving up S&P 500 price informativeness, as well as the decline in the informativeness of smaller assets' prices.

Big Data Alone Should Increase Informativeness of All Prices If investors particularly like processing data about large firms, then perhaps when they have more data processing ability, they direct it towards these large firms. That turns out not to be the case. The next result shows why growth in *data processing alone cannot explain* the facts about price informativeness.

In many cases, after all data processing capacity is allocated, there will be multiple risks with identical $\Lambda(\bar{K}_{jt}, \bar{x}_j)$ weights. That is because the marginal utility of signal precision, Λ_i, is decreasing in the average information precision $\bar{K}_i$. In this case, investors are indifferent about which risk to learn about. When financial data processing efficiency $\bar{B}_t$ rises, more bits are allocated to all the assets in this indifference class.

Lemma 1 ***Technological progress***. *As $\bar{B}_t$ grows, the average investor learns weakly more about every asset j, $\int K_{ij(t+1)}di \geq \int K_{ijt}di$, with strict inequality for all assets that are learned about: $\int K_{ij(t+1)}di > \int K_{ijt}di \;\; \forall j : K_{ijt} > 0$ for some j.*

Figure 9: **Equilibrium Allocation of Data Processing**. Shaded area represents aggregate allocation of data processing. Moving from left to right represents an increase in data processing capacity. More processed data lowers the marginal utility of additional data processing. That causes data on other assets to be processed.



This type of equilibrium is called a "waterfilling" solution (Cover and Thomas, 1991). Figure 9 illustrates how the equilibrium allocation maintains indifference (equal marginal utility) between all assets being learned about. The equilibrium uniquely pins down which risk factors are being learned about in equilibrium, and how much is learned about them, but not which investor learns about which risk factor. Waterfilling arises in other information choice problems, such as Kacperczyk, Nosal, and Stevens (2015).

Lemma 2 ***With more bits, more assets are learned about.*** *If $\bar{x}_i$ is sufficiently large $\forall i$, the set of assets learned about $\mathcal{M}_t$ does not contain all assets, and $\bar{B}_{t+1} - \bar{B}_t$ is sufficiently large, then the set of assets $\mathcal{M}_{t+1}$ learned about in $t+1$ is larger than the set $\mathcal{M}_t$.*

A key force is strategic substitutability in information acquisition, an effect first noted by Grossman and Stiglitz (1980). The more other investors learn about a risk, the more informative prices are and the less valuable it is for other investors to learn about the same risk. If one risk has the highest marginal utility for signal precision, but capacity is high, then many investors will learn about that risk, causing its marginal utility to fall and equalize with the next most valuable risk. With more capacity, the highest two $\Lambda(\bar{K}_{jt}, \bar{x}_j)$'s will be driven down until the equate with the next Λ, and so forth. But when K increases, the marginal utilities of all risks must remain equated. Since learning about any risk reduces its marginal utility, all risks must have more learning about them so that all their marginal utilites remain equal and the economy stays at an optimum.

When big firms get bigger, this alone should reduce the informativeness of small firm prices. The growth in data processing is not the only trend that has changed information processing incentives. At the same time, there has been a change in firm size. It's the growth of large firms that can explain why the informativeness of small firms has not grown. For now, we hold $\bar{B}_t$ fixed and only consider the change in firm size. After we have explored each force separately, we consider their combined effect.

The following result shows that if an asset grows larger, investors process more data about it, on average. But for other assets whose size does not grow, the amount of data processed falls.

Lemma 3 ***Investors value big firm data*** *For K_t and $\bar{x}_j$ sufficiently high, 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 $\partial K_i/\partial \bar{x}_j \leq 0 \; \forall i \neq j$.*

The marginal value of signal precision $\bar{K}_{jt}$ is $\Lambda(\bar{K}_{jt}, \bar{x}_j)$, from (14). Recall that $\partial \Lambda(\bar{K}_{jt}, \bar{x}_j)/\partial \bar{x}_j = \rho^2 \bar{\sigma}_j^2 > 0$. So, larger assets are always more valuable targets for data processing. Next consider the equilibrium data allocation. Equation 14 implies that more capacity is allocated to the larger asset in equilibrium as well.

The fall in data processed about other firms is the consequence of more data about j and a fixed budget for bits of data. If more bits are processed about j, less bits must be processed for some other asset. That decline in bits processed is equally spread across other assets so as to equate the marginal utility of bit processing for all.

Fundamentally, this preference for more data about larger assets comes from the fact that information has increasing returns to scale. A larger asset will be a larger fraction of an average investor's portfolio. One could use all data to learn about a small fraction of one's portfolio value. But that is not as valuable as using data processing to reduce uncertainty about an asset that represents a large fraction of one's portfolio risk and a large fraction of potential profit. The same bit of data can evaluate 1 share or 1000 shares equally well. That makes data that can be applied to many units of asset value – data on large firms – more valuable.

A possible joint effect These results do not prove that financial technology has stagnated. Rather, the facts imply a bound on progress in information processing. For small firm price informativeness to fall, it must be that the rate of increase in data processing is not too large, relative to the growth of large firms.

More data processing capacity increases information processed about small firms. Growth in the size of large firms decreases information processed about smaller firms. Both forces, however, predict a rise in information processing about large firms. Thus, the combination of the two forces can explain both the rise in large firm price informativeness and the stagnation of price informativeness for all other firms.

Of course, this does not prove that these are the two forces responsible for this trend. But the model does offer one coherent way to think through the kinds of forces that might be at work.

4 What Data Should Society Be Processing?

Both the planner and the investor care more about large firms. The investor values data about large firms more, all else equal, because a large firm, by definition, makes up a larger

fraction of the value of an average investor's portfolio. The social planner values data about large firms more because the output of each firm $E[\tilde{d}_{jt}] - l_{jt}$ is scaled by firm size $\bar{x}_j$. In both cases, there are returns to scale in information. But because an investor can rebalance his portfolio to hold more and more of the asset he learns about, whereas a social planner takes the set of firms in the economy as given, the investor has stronger returns to scale than the social planner. Thus, investors prefer to process more data about large firms than what the social planner would prefer.

4.1 A Planner's Problem with Parallel Investor Processing

The planner maximizes the total output by choosing the allocation of investor information acquisition capacity, taking manager optimal effort decision and investor optimal portfolio decision as given.[8] Formally, the planner chooses aggregate signal precisions K_1 and K_2 to maximize

$$\max_{\{\bar{K}_{jt}\}} \quad \sum_j \bar{x}_j \left(\mathbb{E}[\tilde{d}_{jt}] - l_{jt} \right) \tag{18}$$

$$\text{s.t.(16)}, \qquad C_{jt} = \Gamma(\bar{K}_{jt}) \qquad \forall\ j \text{ and} \tag{19}$$

$$\sum_j \bar{K}_{jt} = K_t \tag{20}$$

Note that the constraint on processing power for the planner is linear in signal precision. This is different that the constraint facing the individual investor. It represents the idea of parallel computing and a continuum of investors. The computing of different investors is done independently. If each investor can process a total of b_I bits, which results in a signal of precision k_I, then producing a signal with double that precision requires two investors, each processing b_I bits, each producing a conditionally independent signal of precision k_I. Bayes law tells us that if we combine two conditionally independent, normal signals, each with precision k_I, the total precision of the optimally combined signals is $2k_I$. So, double the precision requires double the resources, implying a linear constraint on signal precision.

The first order condition of this problem with respect to $\bar{K}_{jt}$ is

[8] Assume investors have sufficient wealth that their marginal utility is vanishingly small and drops out of planner objective.

$$\bar{x}_j \Gamma^{-2}(\bar{K}_{jt})\Gamma'(\bar{K}_{jt})\left((g')^{-1}\right)'(\bar{K}_{jt})\Big(g'(\bar{K}_{jt})-1\Big)=\mu \tag{21}$$

where μ is the Lagrange multiplier, or shadow cost, of one additional unit of aggregate signal precision.

In general, equilibrium outcomes and constraint efficient allocation are different. We can see this from the fact that the solutions to equations 15 and 21 do not coincide. But why are individual and social choices different and what are the economics behind this difference?

If we substitute $\mathbb{E}[\tilde{d}_{jt}] = g(l_{jt}) = l_{jt}^{\phi}$ and then use (16) to substitute for l, we get a simplified planner problem:

$$\begin{aligned} \max_{\{\bar{K}_{jt}\}} \quad & \sum_j \bar{x}_j \Big(\big(\phi\Gamma(\bar{K}_{jt})\big)^{\frac{\phi}{1-\phi}} - \big(\phi\Gamma(\bar{K}_{jt})\big)^{\frac{1}{1-\phi}} \Big) \\ \text{s.t.} \quad & \sum_j \bar{K}_{jt} = K \end{aligned}$$

Merging the first order condition of the planner with respect to any two assets j, j' we get

$$\frac{\big(1-1/\Gamma(\bar{K}_{jt})\big)\,\Gamma(\bar{K}_{jt})^{\frac{\phi}{1-\phi}}\Gamma'(\bar{K}_{jt})}{\big(1-1/\Gamma(\bar{K}_{j't})\big)\,\Gamma(\bar{K}_{j't})^{\frac{\phi}{1-\phi}}\Gamma'(\bar{K}_{j't})} = \frac{\bar{x}_{j'}}{\bar{x}_j} \tag{22}$$

Let F represent the marginal social value of an additional unit of data precision, per share of the asset,

$$F(\bar{K}_{jt}) = \big(1-\Gamma^{-1}(\bar{K}_{jt})\big)\,\Gamma(\bar{K}_{jt})^{\frac{\phi}{1-\phi}}\Gamma'(\bar{K}_{jt}).$$

Then with two assets, we can express the social optimum simply as $F(\bar{K}_{2t})/F(\bar{K}_{1t}) = \bar{x}_1/\bar{x}_2$.

4.2 Why the Social Optimum Involves Less Data on Large Firms

For the investor, the potential profits from learning more and more precise information are unbounded. But for a social planner, the gains to information from added efficiency are bounded. From differentiating (17), we learn that $\frac{\partial C_{jt}}{\partial \bar{K}_{jt}} > 0$ and $\lim_{\bar{K}_{jt}\to\infty} C_{jt} = 1$. Thus, an

infinite amount of data can only possibly make price informativeness 1 at most. This offers finite social welfare gains.

Lemma 4 ***The improvement in price informativeness from additional data processing exhibits diminishing returns.*** *If K_t is sufficiently large, then $\partial^2 C_{jt}/\partial \bar{K}_{jt}^2 < 0$.*

To ensure that the second order condition of the planner problem is satisfied, it must be that $F'(K) < 0$, which holds when K is sufficiently large. Inspecting the objective function of the planner, it is easy to verify that the planner allocates more capacity to the larger asset, proportional to its marginal social value, its supply $\bar{x}_i$. This observation is also verified in equation 22 using the second order condition. The difference is governed by concavity of the production function.

The fact that the result rests on a sufficiently high level of data processing explains why this phenomenon of informative large firm prices has grown over time. When K was small, the social planner valued large firm data more than the investor. As K grew larger, the stronger increasing returns to data in large firms for investors kicked in, and large firm prices became more informative.

Let $\{\bar{K}_{jt}^{sp}\}_j$ and $\{\bar{K}_{jt}^{eq}\}_j$ denote the solution to the constraint planner and equilibrium. With two assets, the following two equations fully characterize the two solutions when both assets are learned about[9]

$$\bar{x}_1 F(\bar{K}_{1t}^{sp}) - \bar{x}_2 F(K_t - \bar{K}_{1t}^{sp}) = 0$$
$$\Lambda(\bar{K}_{1t}^{eq}, \bar{x}_1) - \Lambda(K_t - \bar{K}_{1t}^{eq}, \bar{x}_2) = 0$$

where Λ is defined in (14).

It is straightforward to verify that $\forall\ (\tau_0, \tau_x, \frac{\bar{x}_1}{\bar{x}_2}, \phi);\ 1 < \frac{\bar{x}_1}{\bar{x}_2} < x_{\max},\ 0 < \phi < 1$, if $\rho > \bar{\rho}$ then $\bar{K}_{1t}^{eq} > \bar{K}_{1t}^{sp}$ and $\bar{K}_{2t}^{eq} < \bar{K}_{2t}^{sp}$. In other words, in equilibrium investors learn too much about the larger firm and the smaller firm remains under unexplored.

[9]Note that in both equilibrium and planner problem it might be that only one asset is learned about, when $\bar{x}_1 >> \bar{x}_2$. Consider only the set of parameters that this does not happen.

Why does equilibrium feature a misallocation of resources away from the smaller risk toward the larger risk? Although it is true that both the constraint social planner and individual investors care about the larger asset more, the investor preferences are more extreme since information has increasing return to scale at the individual level, but only constant return to scale at the aggregate level.

4.3 How Inefficient? A Numerical Example

To get a sense of how much data processing might depart from the social optimum, we consider a two-firm numerical example. We explore three possible long-run trends. The first is an increase in the efficiency of data processing; the second is an increase in the size of the larger firm, and the third trend is a joint increase in data processing and large firm size.

All three exercises use a common set of parameters. We choose the exponent on effort in the manager's production function to be $\phi = 0.4$; risk aversion is $\rho = 4$; the inverse variance of the dividend payoff is $\tau_0 = 1$; the inverse variance of asset supply shock is $\tau_x = 3$. To think about firm size effects, we need a large and small firm. So, the average asset supply for asset 1 is set to be larger so that asset 1 is the large asset and asset 2 is the smaller firm. For the exercise where we vary only data processing, the average asset supply is $\bar{x}_1 = 2$ for asset 1 and $\bar{x}_2 = 1$ for asset 2. The total capacity grows at a constant rate each period, starting at $K_t = 8$ and ending at $K_t = 16$. For the exercise where only firm size grows, capacity is fixed at $K_t = 10$ and the large firm size $\bar{x}_2$ grows from 1 to 2. When both grow, K_t starts at 8 and $\bar{x}_1$ starts at $\bar{x}_2$, and both grow by 0.1 each period. We did a handful of robustness checks by varying parameters within an order of magnitude and found qualitatively similar results.

Comparing Figures 10 and 11 reveals that while growth in data processing creates very little social inefficiency, growth in the size of the largest firm opens up a large gap between the competitive outcome and the social optimum. As the size of the largest firm grows, it is socially efficient to learn more about it. It is economically more important, so pricing the firm correctly is socially more valuable. However, the economy processes too much data about large firms. The forces in Lemma 2, which make investors profit more from large firm data, all else equal, are stronger than the social benefit to large firm data processing, which

Figure 10: **Data Tech Growth: Competitive vs Optimal Data Processing**. Total information processing capacity starts at $\kappa = 8$ and grows to 16. Other parameters are constant. Asset supply is $\bar{x}_1 = 2$ for asset 1 and $\bar{x}_2 = 1$ for asset 2.

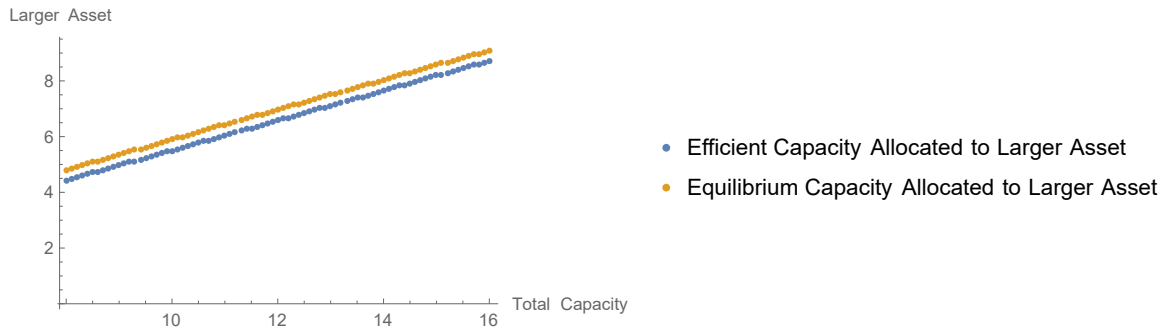


Figure 11: **Large Firms Grow: Competitive vs Optimal Data Processing**. Asset supply for the large firm grows from $\bar{x}_2 = 1$ to 2. Other parameters are constant. Total information processing capacity is fixed at $\kappa = 10$.

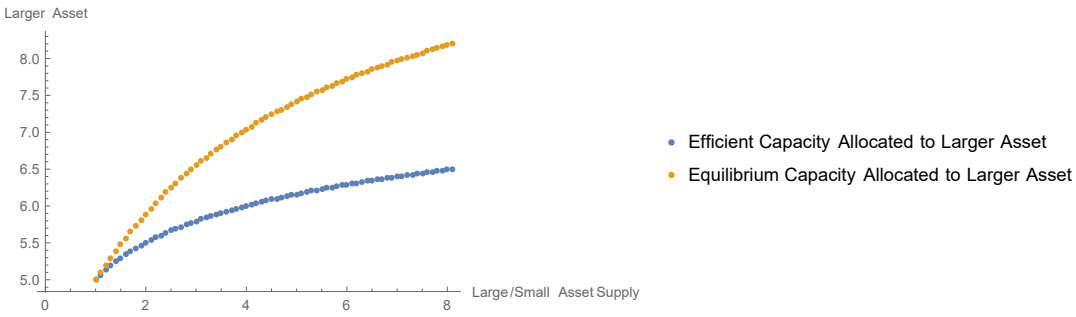
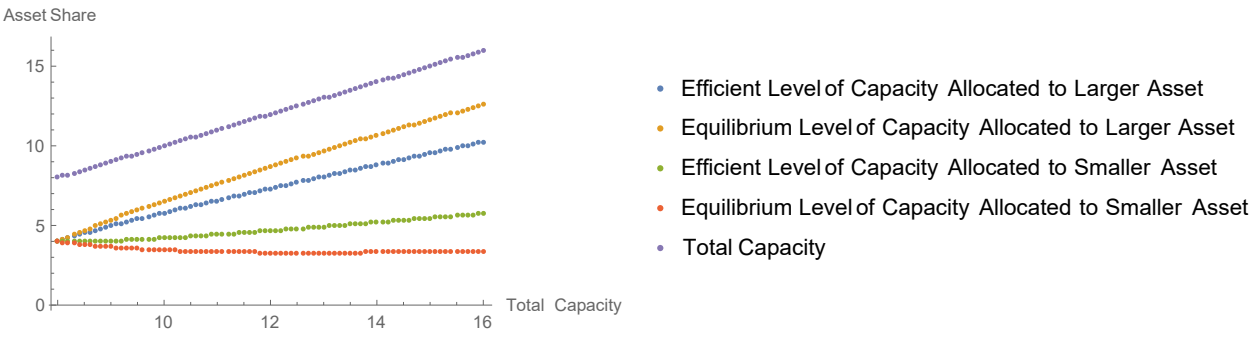


Figure 12: **Data Tech and Firm Size Both Increase: Competitive vs Optimal Data Processing**. Data processing capacity starts at $K = 8$, large firm size x_2 starts at 1. Both grow by 0.1 each period. Other parameters are constant.



exhibits diminishing returns (Lemma 4). Thus, as the large firm gets larger and larger, data processing decisions become more and more inefficient.

Finally Figure 12 shows a case where both data processing and large firm size grow. In this example, data processing growth is slow enough that the amount of data processing about the small firm declines. When data processing about small firms declines, the informativeness of small firm prices declines as well. Thus, this numerical example shows how the model can speak to the decline of price informativeness among non-S&P 500 firms, observed in the data.

Notice, however, that the social planner would want data processing about small firms to rise. Of course, efficient data processing about the large firm would rise by more, but the planner does not want the informativeness of small firm prices to fall here. The lesson from these results is that the data are compatible with a world where the total efficiency (K) of the financial sector is growing. But that growth is limited. In particular, financial efficiency growth is bounded by the rate of increase of large (S&P 500) firm size.

4.4 What if Investors' Computing Could be Integrated?

The reason that the social planner's problem features a linear constraint is that each investor in the economy produces a conditionally independent signal. They process data independently. When different processors work simultaneously, but independently on a problem, that is called parallel computing. For a given investor, the constraint on computing is not linear because optimal data processing is not parallel. A single processor can accomplish more than two processors, each with half the power, because its processing is integrated. With integrated computing, twice as many bits can transmit more than double the precision of signal.

This raises the question, what if economy-wide computing became integrated? Instead of each investor processing their data in parallel, what if all computing were done on a common processor? This idea of futuristic cloud computing both is a speculation about future technology, but also a way of breaking down the difference between the social planner and decentralized problem into the technological differences between integrated and parallel computing, and the payoff externalities internalized by the planner. This formulation of the

problem gives the planner and the individual the same computing constraints and focuses only on the payoff externalities.

$$\begin{aligned} \max_{\{\bar{K}_{jt}\}} \quad & \sum_j \bar{x}_j \Big(\big(\phi\Gamma(\bar{K}_{jt})\big)^{\frac{\phi}{1-\phi}} - \big(\phi\Gamma(\bar{K}_{jt})\big)^{\frac{1}{1-\phi}} \Big) \\ \text{s.t.} \quad & \sum_j ln(1+\tau_0^{-1}\bar{K}_{jt}) = \bar{B}_t \end{aligned}$$

and, of course $\bar{K}_{jt} \geq 0$. The first order condition of the planner is the same as before, except that the Lagrange multiplier is multiplied by the derivative of $ln(1+\tau_0^{-1}\bar{K}_{jt})$, which is $\tau_0^{-1}/(1+\tau_0^{-1}\bar{K}_{jt})$ or alternatively $(\tau_0+\bar{K}_{jt})^{-1}$. Working through the same steps as before for the two assets, we can express the social optimum simply as

$$\frac{F(\bar{K}_{2t})(\tau_0+\bar{K}_{2t})}{F(\bar{K}_{1t})(\tau_0+\bar{K}_{1t})} = \frac{\bar{x}_1}{\bar{x}_2}. \tag{23}$$

Notice that this solution is as if the marginal social value of data is more increasing, or less decreasing than before. In other words, integrated computing created more increasing returns to processing the same type of data at the aggregate level.

Does this mean that the social planner has increasing returns and will only want to process data on large firms? Probably not. It depends on the parameter values. Results on precise bounds are work in progress. But this does suggest that a future shift to more integrated computing methods would make more concentrated computing more desirable and bring the social optimum and decentralized equilibrium closer.

5 Conclusion

Is the financial sector adding social value? One of the most important functions of financial markets is to produce information that is used to allocate capital and effort and put it to its most productive uses. One indicator that a financial market is effectively performing these tasks is that its prices are good predictors of future firm value (price informativeness).

Price informativeness is trending upward, but only for large, old firms. For others, there

is a stagnation, or even slight decline. Thus, the rise in price informativeness of S&P 500 firms appears to be driven by a composition effect, as the firms in the S&P 500 become older. These older, larger firms have more informative financial prices. But the informativeness of a firm of a given size has not increased perceptibly.

To understand these facts, we consider a model with two long-run trends. On the one hand, there is an increase in the efficiency of data processing over time. That is important to consider because when people talk about the financial sector becoming more efficient, often the is associated with greater information efficiency, suggesting the more information is being discovered, processed and aggregated. While such a trend can indeed explain why old, large firms have more informative prices, it cannot explain the stagnation of small firm price informativeness. We show that a rise in data processing, alone, should add information content to all prices.

The second trend we consider is the well-documented fact that large firms are growing larger. The trend holds also for the publicly-listed firms in our sample. Our model clarifies exactly why it is that such large firms are more valuable to process data on. As they grow bigger, investors' optimal allocation of data processing shifts towards these growing firms. As more data is processed and used by investors to trade, the price informativeness of such firms rises. The combination of the two forces can explain the joint evolution of large and small firm price informativeness.

This trend is not consistent with social welfare maximization. Investors get greater private returns from processing data on large firms than the social return. This is because an individual who knows more about an asset can buy more of it and hold more of it. This creates a private advantage to having information about assets one might hold lots of. But at the economy-wide level, more information does not make more of a firm appear. So the returns to processing data about large firms are still higher than for small firms, but not as high as the private returns to data for large firms.

Finally, we explore the difference between the social optimum with parallel computing and integrated computing. If in the future, all computers were integrated, the social optimum would then feature more concentrated data processing and the social and private optimal data processing choices might well converge.

References

Bai, J., T. Philippon, and A. Savov (2013): "Have Financial Markets Become More Informative?," NBER Working Paper 19728.

Chordia, T., C. Green, and B. Kottimukkalur (2016): "Rent Seeking by Low Latency Traders: Evidence from Trading on Macroeconomic Announcements," Working Paper, Emory University.

Cover, T., and J. Thomas (1991): *Elements of information theory*. John Wiley and Sons, New York, New York, first edn.

Crouzet, N., I. Dew-Becker, and C. Nathanson (2016): "A Model of Multi-Frequency Trade," Northwestern University Working Paper.

David, J., H. Hopenhayn, and V. Venkateswaran (2016): "Information frictions, Misallocation and Aggregate Productivity," *Quarterly Journal of Economics*, 131 (2), 943–1005.

Davila, E., and C. Parlatore (2016): "Trading Costs and Informational Efficiency," NYU Working Paper.

Dugast, J., and T. Foucault (2016): "Data Abundance and Asset Price Informativeness," Working Paper, HEC.

Durnev, A., R. Morck, and B. Yeung (2004): "Value-enhancing capital budgeting and firm-specific stock return variation," *The Journal of Finance*, 59(1), 65–105.

Edmans, A., S. Jayaraman, and J. Schneemeier (2016): "The source of information in prices and investment-price sensitivity," .

Farboodi, M., and L. Veldkamp (2017): "Long Run Growth of Financial Technology," Discussion paper, National Bureau of Economic Research.

Grossman, S., and J. Stiglitz (1980): "On the impossibility of informationally efficient markets," *American Economic Review*, 70(3), 393–408.

Hsieh, C., and P. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124(4), 1403–1448.

Kacperczyk, M., J. Nosal, and L. Stevens (2015): "Investor Sophistication and Capital Income Inequality," Imperial College Working Paper.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016): "A Rational Theory of Mutual Funds' Attention Allocation," *Econometrica*, 84(2), 571–626.

Kyle, A., and J. Lee (2017): "Toward a Fully Continuous Exchange," SSRN Working Paper.

Maćkowiak, B., and M. Wiederholt (2009): "Optimal sticky prices under rational inattention," *American Economic Review*, 99 (3), 769–803.

Manzano, C., and X. Vives (2010): "Public and Private Learning from Prices, Strategic Substitutability and Complementarity, and Equilibrium Multiplicity," CEPR Discussion Papers 7949, C.E.P.R. Discussion Papers.

Mondria, J. (2010): "Portfolio choice, attention allocation, and price comovement," *Journal of Economic Theory*, 145, 1837–1864.

Mondria, J., T. Wu, and Y. Zhang (2010): "The determinants of international investment and attention allocation: Using internet search query data," *Journal of International Economics*, 82 (1), 85–95.

Pástor, L., and R. F. Stambaugh (2012): "On the size of the active management industry," *Journal of Political Economy*, 120, 740–781.

Ranco, G., D. Aleksovski, G. Caldarelli, M. Grcar, and I. Mozetic (2015): "The Effects of Twitter Sentiment on Stock Price Returns," PLOS working paper.

Restuccia, D., and R. Rogerson (2013): "Misallocation and Productivity," *Review of Economic Dynamics*, 16, 1–10.

Sims, C. (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50(3), 665–90.

Stambaugh, R. (2014): "Presidential Address: Investment Noise and Trends," *Journal of Finance*, 69 (4), 1415–1453.

Van Nieuwerburgh, S., and L. Veldkamp (2009): "Information immobility and the home bias puzzle," *Journal of Finance*, 64 (3), 1187–1215.

A Appendix

A.1 Model Solution Details

Price Coefficients: Equating supply and demand from (11), we get

$$\int \frac{1}{\rho}\hat{\Sigma}^{-1}\left(E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r\right) = \bar{x} + x_t \tag{24}$$

where $\hat{\Sigma}^{-1} \equiv Var[\tilde{d}_t|\mathcal{I}_{it}]$. Rearranging this expression yields $p = 1/2[E[\tilde{d}_t|\mathcal{I}_{it}] - \rho(\bar{x} + x_t)]$. If we then substitute in the conditional expectation from (10), and use the defintion of the price signal $\eta_p = C_t^{-1}(p_t - A_t)$, to obtain

$$p = \frac{1}{r}\hat{\Sigma}\left[\Sigma^{-1}\mu + \int K_{it}\eta_{it}di + \Sigma_p^{-1}C_t^{-1}(p_t - A_t) - \rho(\bar{x} + x_t)\right] \tag{25}$$

Notice the price p on the left and right side of the equation. The term on the right is from the fact that agents use price as a signal. Next, we collect terms in p. We also use the fact that since signals are unbiased, irrespective of precision, $\int K_{it}\eta_{it}di = \bar{K}_t\tilde{d}_t$. The resulting equation is (12), where

$$A = \frac{1}{r}\hat{\Sigma}(1 - \frac{1}{r}\hat{\Sigma}\Sigma_p^{-1}C^{-1})^{-1}(\Sigma^{-1}\mu - \Sigma_p^{-1}C^{-1}A - \rho\bar{x}) \tag{26}$$

$$C = 1 - \tau_{0j}Var[g(l_{jt}) + \tilde{y}_{jt}|\mathcal{I}_{ijt}] \tag{27}$$

$$D = -C\frac{\rho}{\bar{K}_t} \tag{28}$$

Price information is the signal about the payoff vector $\tilde{d}_t$ contained in prices. The transformation of the price vector p_t that yields an unbiased signal about $\tilde{d}_t$ is $\eta_p \equiv C^{-1}(p_t r - A)$. The signal noise in prices is $\varepsilon_p = C^{-1}Dx$. Since we assume $x \sim N(0, \sigma_x I)$, the price noise is distributed $\varepsilon_p \sim N(0, \Sigma_p)$, where $\Sigma_p \equiv \sigma_x C^{-1}DD'C^{-1'}$. Substituting in the coefficients C and D shows that the signal precision of prices is $\Sigma_p^{-1} = \bar{K}_t\bar{K}_t/(\rho^2\sigma_x)$ is a diagonal matrix and $\bar{K}_t \equiv \int K_{it}di$. The jth diagonal element of $\bar{K}_t$ is the average capacity allocated to each asset j at date t.

Computing ex-ante expected utility: Substitute optimal risky asset holdings from equation (11) into the first-period objective function to get: $U_{1j} = rW_0 + \frac{1}{2}E_1\left[(E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r)\hat{\Sigma}^{-1}(E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r)\right]$. Note that the expected excess return $(E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r)$ depends on signals and prices, both of which are unknown at time 1. Because asset prices are linear functions of normally distributed shocks, $E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r$, is normally distributed as well. Thus, $(E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r)\hat{\Sigma}^{-1}(E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r)$ is a non-central χ^2-distributed variable. Computing its mean yields:

$$U_{1j} = rW_0 + \frac{1}{2}trace(\hat{\Sigma}^{-1}V_1[E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r]) + \frac{1}{2}E_1[E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r]'\hat{\Sigma}^{-1}E_1[E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r]. \tag{29}$$

Note that in expected utility (29), the choice variables K_{ij} enter only through the posterior

variance $\hat{\Sigma}$ and through $V_1[E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r] = V_1[\tilde{d}_t - p_t r] - \hat{\Sigma}$. Since there is a continuum of investors, and since $V_1[\tilde{d}_t - p_t r]$ and $E_1[E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r]$ depend only on parameters and on aggregate information choices, each investor takes them as given.

Since $\hat{\Sigma}^{-1}$ and $V_1[E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r]$ are both diagonal matrices and $E_1[E[\tilde{d}_t|\mathcal{I}_{it}] - p_t r]$ is a vector, the last two terms of (29) are weighted sums of the diagonal elements of $\hat{\Sigma}^{-1}$. Thus, (29) can be rewritten as $U_i = rW_0 + \sum_j \Lambda_j \hat{\Sigma}^{-1}(j,j) - n/2$, for positive coefficients Λ_j. Since rW_0 is a constant and $\hat{\Sigma}^{-1}(j,j) = \Sigma^{-1}(j,j) + \Sigma_p^{-1}(j,j) + K_{ij}$, the information choice problem is (13).

A.2 Proofs

Proof of Lemma 1. From step 4 of the model solution, we know that when there is a unique maximum Λ_{lt} the optimal information choice is $K_{ilt} = K = \tau_0(\exp \bar{B}_t - 1)$ if $\Lambda_{lt} = \max_j \Lambda_{jt}$, and $K_{ijt} = 0$, otherwise. If multiple risks achieve the same maximum Λ_l then all attention will be allocated amongst those risks, but each investor would learn about one single risk.

First, we show that value of learning about asset j falls as the aggregate capacity devoted to studying it increases: $\partial\Lambda(\bar{K}_{jt}, \bar{x}_j)/\partial\bar{K}_{jt} < 0$. This is the same strategic substitutability in information as in Grossman and Stiglitz (1980). The solution for Λ_j is given by (14). It is clearly increasing in $\bar{K}_{jt}$ directly. But there is also an indirect negative effect through $\bar{\sigma}_j$. Recall that by Bayes' Law, the average posterior precision is $\bar{\sigma}_j^{-1} = \sigma_j^{-1} + \sigma_{pj}^{-1} + \bar{K}_{jt}$. Thus, $\frac{\partial\bar{\sigma}_j}{\partial\bar{K}_{jt}} < 0$. To sign the net effect, it is helpful to rewrite Λ_j as

$$\Lambda_j = \bar{\sigma}_j^2 \left[\bar{\sigma}_j^{-1} + \rho^2(1/\tau_x + \bar{x}_j^2) + \bar{K}_{jt}\right]$$

Substituting in $\bar{\sigma}_j^{-1} = \sigma_j^{-1} + \sigma_{pj}^{-1} + \bar{K}_{jt}$, we get

$$\Lambda_j = \frac{\sigma_j^{-1} + \sigma_{pj}^{-1} + \rho^2(1/\tau_x + 2\bar{x}_j^2) + \bar{K}_{jt}}{(\sigma_j^{-1} + \sigma_{pj}^{-1} + \bar{K}_{jt})^2}$$

Finally, the partial derivative with respect to $\bar{K}_{jt}$ is

$$\frac{\partial\Lambda_j}{\partial\bar{K}_{jt}} = \frac{2(\sigma_j^{-1} + \sigma_{pj}^{-1} + \bar{K}_{jt}) - 2(\sigma_j^{-1} + \sigma_{pj}^{-1} + \rho^2(1/\tau_x + 2\bar{x}_j^2) + \bar{K}_{jt})}{(\sigma_j^{-1} + \sigma_{pj}^{-1} + \bar{K}_{jt})^3}$$

$$= \frac{-2\rho^2(1/\tau_x + 2\bar{x}_j^2) - 2\bar{K}_{jt})}{(\sigma_j^{-1} + \sigma_{pj}^{-1} + \bar{K}_{jt})^3} < 0$$

Since the numerator is all terms that can only be negative and the denominator is a sum of precisions, that can only be positive, the sign is negative. This proves that Λ_j is decreasing in $\bar{K}_{jt}$.

Next, assume that in the equilibrium information allocation, the claim of the lemma does not hold. Either, (1) as capacity $\bar{B}_t$ rises, there exists an asset that investors learn strictly less about; or (2) there exists an asset which is learned about, and the average investor learns weakly less about

it as $\bar{B}_t$ grows.First, assume the latter is true, and let $\mathcal{M}_t$ denote the set of assets learned about at time t. We have $\bar{B}_{t+1} > \bar{B}_t$, $\hat{j} \in \mathcal{M}_t$, and $\int K_{i\hat{j}(t+1)}di \leq \int K_{i\hat{j}t}$. There are two cases:

Case 1: $\exists j \in \mathcal{M}_t$ such that $\int K_{ij(t+1)}di > \int K_{ijt}$. Since $j, \hat{j} \in \mathcal{M}_t$, $\Lambda_{jt} = \Lambda_{\hat{j}t}$. Moreover, by the above strategic substitutability argument, $\Lambda_{j(t+1)} < \Lambda_{jt}$ and $\Lambda_{\hat{j}(t+1)} \geq \Lambda_{\hat{j}t}$. So investors who have learned about j have an incentive to switch and learn about $\hat{j}$ instead, and the proposed allocation cannot be an equilibrium.

Case 2: $\nexists j \in \mathcal{M}_t$ such that $\int K_{ij(t+1)}di > \int K_{ijt}$. Since all the capacity is exhausted in equilibrium (linear objective subject to concave constraint), there must be at least one other asset, l, which is learned about at $t+1$. Let $l^* \in argmax_{j \notin \mathcal{M}_t} \Lambda(\bar{K}_{j(t+1)}, \bar{x}_j)$. Since $l^* \notin \mathcal{M}_t$, $K_{l^*t} = 0$, and $\Lambda(0, \bar{x}_{l^*}) < \Lambda(\bar{K}_{jt}, \bar{x}_j)$, $\forall\ j \in \mathcal{M}_t$. Since Λ_{jt} is a continuous function of $\bar{K}_{jt}$, there is a reallocation of a subset of investors who have learned about l^* to some asset $j \in \mathcal{M}_t$ which strictly increases the utility of both the reallocated investors and those who still learn about l^*. More specifically, two cases can be considered separately.

Let $j^* \in argmax j \in \mathcal{M}_t \Lambda(\bar{K}_{j(t+1)}, \bar{x}_j)$.

1. If $\Lambda_{l^*(t+1)} < \Lambda_{j^*(t+1)}$, then every investor learning about l^* has an incentive to switch and learn about j^*, a contradiction.

2. If $\Lambda_{l^*(t+1)} \geq \Lambda_{j^*(t+1)}$, then since $\Lambda_{l^*t} = 0 < \Lambda_{j^*t}$, by continuity of Λ function there is a reallocation of investors from l^* to j^* which makes every investor better off, so the original allocation cannot be an equilibrium.

Now assume the former part of the claim is violated. If $j \in \mathcal{M}$, the second part already proves a contradiction. If $j \not\in \mathcal{M}$, then $\bar{K}_{jt} = 0$. So a violation requires $\bar{K}_{j(t+1)} < 0$ which is not possible.

Since each investor chooses to learn exactly about one asset, higher $\bar{B}_t$ maps one to one to higher K_{jit}, and higher $\bar{K}_{jt}$, which completes the proof.

■

Proof of Lemma 3. As in the previous lemma, we know that when there is a unique maximum Λ_{lt} the optimal information choice is $K_{ilt} = K = \tau_0(\exp \bar{B}_t - 1)$ if $\Lambda_{lt} = \max_j \Lambda_{jt}$, and $K_{jlt} = 0$, otherwise. If multiple risks achieve the same maximum Λ_l then all attention will be allocated amongst those risks, but each investor would learn about one single risk. Therefore, there are three cases to consider.

Case 1: Λ_{lt} is the unique maximum Λ_{jt}. Holding attention allocations constant, a marginal increase in $\bar{x}_l$ will cause Λ_{lt} to increase:

$$\partial \Lambda(\bar{K}_{jt}, \bar{x}_j)/\partial \bar{x} = \rho^2 \bar{\sigma}_j^2 > 0.$$

The marginal increase in $\bar{x}_l$ will not effect $\Lambda_{l't}$ for $l' \neq l$. It follows that after the increase in $\bar{x}_l$, Λ_{lt} will still be the unique maximum Λ_{jt}. Therefore, in the new equilibrium, attention allocation is unchanged.

Case 2: Prior to the increase in $\bar{x}_l$, multiple risks, including risk l, attain the maximum Λ_{jt}, with $\mathcal{M}_t$ denoting the set of such risks. If $\bar{x}_l$ marginally increases and we held attention allocations fixed, then Λ_{lt} would be the unique maximum Λ_{jt}. If Λ_{lt} is the unique maximum, then more investors have to learn about risk l, $\bar{K}_{lt}$ increases, which implies fewer investors learn about any other risk $l \in \mathcal{M}_t \setminus l$, $\bar{K}_{l't}$ decreases. However, lemma A.2 shows that an increase in $\bar{K}_{lt}$ would decrease Λ_{lt}. Recall that $\bar{K}_{lt} = K_{ilt}$ for all investors who learn about asset l. This effect works to partially offset the initial increase in Λ_{lt} as fewer investors will have an incentive to learn about l. In the rest of the proof, we construct the new equilibrium attention allocation, following an initial increase Λ_{lt} and show that even though the attention reallocation works to reduce Λ_{lt}, the net effect is a larger $\bar{K}_{lt}$.

This solution to this type of convex problem is referred to as a "waterfilling" solution in the information theory literature (See textbook by Cover and Thomas (1991)). To construct a new equilibrium, we reallocate attention from risk $l' \in \mathcal{M} \setminus l$ to risk l (increasing the number of investors who learn about l and as a result $\bar{K}_{lt}$, decreasing the number of investors who learn about l' and as a result $\bar{K}_{l't}$). This decreases Λ_{lt} and increases $\Lambda_{l't}$. We continue to reallocate attention from all risks $l' \in \mathcal{M} \setminus l$ to risk l in such a way that $\Lambda_{l't} = \Lambda_{l''t}$ for all $l', l'' \in \mathcal{M} \setminus l$ is maintained. We do this until either (i) all attention has been allocated to risk l or (ii) $\Lambda_{lt} = \Lambda_{l't}$ for all $l' \in \mathcal{M} \setminus l$. Note that in the new equilibrium Λ_{lt} will be larger than before and the new equilibrium $\bar{K}_{lt}$ will be larger than before, while $\bar{K}_{l't}$, $l' \in \mathcal{M} \setminus l$ will be smaller than before.

Case 3: Prior to the increase in $\bar{x}_l$, $\Lambda_{lt} < \Lambda_{l't}$ for some $l' \neq l$. Because Λ_{lt} is a continuous function of $\bar{x}_l$, a marginal increase in $\bar{x}_l$, will only change Λ_{lt} marginally. Because Λ_{lt} is discretely less than $\Lambda_{l't}$, the ranking of the Λ_{it}'s will not change and the new equilibrium will maintain the same attention allocation.

In cases one and three $\bar{K}_{lt}$ does not change in response to a marginal increase in $\bar{x}_l$. In case two $\bar{K}_{lt}$ is strictly increasing in $\bar{x}_l$. Therefore, $\bar{K}_{lt}$ is weakly increasing in $\bar{x}_l$.

■

Proof of Lemma 2.

To show: If $\bar{x}_i$ is sufficiently large $\forall i$, the set of assets learned about $\mathcal{M}_t$ does not contain all assets, and $\bar{B}_{t+1} - \bar{B}_t$ is sufficiently large, then the set of assets $\mathcal{M}_{t+1}$ learned about in $t+1$ is larger than the set $\mathcal{M}_t$.

Suppose not. Then, there would be a unique maximum set $\Lambda_j \forall j \epsilon \mathcal{M}_t$ that is non-increasing, no matter how large $\bar{B}_{t+1}$ is. Since there is a unique maximum, the equilibrium solution dictates that all information capacity is used to study this set of risks. Thus the averge precision of information, $\bar{K}_{jt} \equiv \int K_{ijt} di$ becomes arbitrarily large $\forall j \epsilon \mathcal{M}_t$.

However, the value of learning about asset j falls as the aggregate capacity devoted to studying it increases: $\partial \Lambda_j / \partial \bar{K}_{jt} < 0$. Furthermore, as the supply of the risk factor $\bar{x}_j$ becomes large,

$\partial\Lambda_j/\partial\bar{K}_{jt}$ becomes an arbitrarily large negative number. Thus, for a sufficiently large $\bar{x}_j$, there exists a $\underline{K}$ such that if $\bar{K}_{jt} = \underline{K}$, then $\Lambda_j < \Lambda_{j'}$ for some other risk j'. But then, Λ_i is not a unique maximum in the set of $\{\Lambda_l\}_{l=1}^N$, which is a contradiction. Thus the set of assets learned about $\mathcal{M}_{t+1}$ must grow. ■

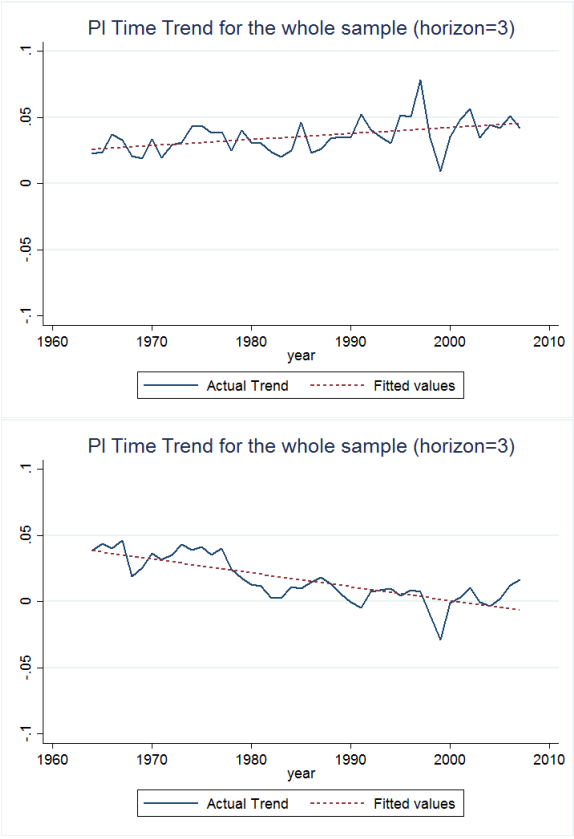
Proof of Lemma 4. Differentiating (17) a second time,

$$\frac{\partial^2 C_j}{\partial K_j^2} = -\frac{2\rho^2\tau_{0j}\left(3K_j^2\tau_{xj}^2 + \rho^2\tau_{xj}(3K_j - \tau_{0j}) + \rho^4\right)}{\left(K_j^2\tau_{xj} + \rho^2(K_j + \tau_{0j})\right)^3}$$

So as long as $K_j \geq \frac{\tau_{0j}}{3}$, the numerator is positive and thus the second derivative is negative, which completes the proof. □ ■

A.3 Robustness of Empirical Results

Figure 13: **Price Informativeness is Rising for S&P 500 Firms and Falling for All Public Firms**. Price informativeness is the ability to forecast future earnings (eqn 2). The sample is all firms in the S&P 500 at that date. The future earnings are measured here at a 3-year, instead of a 5-year horizon.



For S&P 500 with $k = 3$, the mean is at 3.4, the standard deviation is at 1.12, and the interquartile range is at 1.24. So over the period, (since the time trend is normalized between 0 and 1) it means that price has increased by 1.57 times the s.d., by 52% relative to the mean and by by almost 50% relative to the interquartile range. For non-S&P 500 firms with $k = 3$, the mean is 2.4, standard deviation is 1.2 and the interquartile range is 1.8. So price informativeness dropped by 127% relative to the mean, 2.6 times relative to the standard deviation and 1.7 times relative to the interquartile range.