Recovering Price Informativeness from "Nonfundamental" Shocks

Julien Cujean* Samuel Jaeger†

August 29, 2024

Abstract

We measure price informativeness (PI) based on price recovery from supply shocks using mutual fund fire sales. Incomplete observability of shocks suggests a rational-expectations view: investors learn which of fundamentals or supply drive price pressures they see together with their private information. In equilibrium, PI flows with the square of the private information flow, a deterministic, nondecreasing but otherwise arbitrary function. The main result is a formula that recovers this function from the quadratic variation of returns on shocked firms using intraday data. This procedure produces *curves* of PI for each shock. Since the late 90s prices have revealed less information more slowly (PI's slope and curvature) while fundamental uncertainty has declined (level). Crises coincide with sudden stops and equally fast rebounds in the PI flow when they end. Size and liquidity largely determine the shape of PI across firms, yet are of vanishing importance over the last decade.

[0]We thank Julien Hugonnier, Boris Nikolov, Norman Schuerhoff, Philip Valta, Marco Zanotti and seminar participants at the University of Lausanne, EPFL, the University of Neuchatel and the University of Bern.

*CEPR and the University of Berne, Institute for Financial Management, Engehaldenstrasse 4, 3012 Bern, Switzerland, juliencujean.com, julien.cujean@unibe.ch.

†University of Berne, Institute for Financial Management, Engehaldenstrasse 4, 3012 Bern, Switzerland, samuel.jaeger@unibe.ch.

1 Introduction

[In a society where information is dispersed among many people] we must look at the price system as a mechanism for communicating information if we want to understand its real function.

—F. Hayek (1945)

The classical formalization of this idea is that prices are signals of the form "value plus noise" (e.g., Grossman and Stiglitz (1980)). Measuring informativeness of prices consists then in mapping variation in prices into variation in "value" or "noise". The issue is that the two are not observable separately. The literature *1.* proxies for the "value" element of the price signal with future earnings, which it then *2.* regresses on prices (e.g., Bai, Philippon, and Savov (2016)). Yet, this expedient raises other, practical problems: "earnings are statistics conceived by accountants which are supposed to provide an indicator of how well a company is doing, and there is a great deal of latitude for the definition of earnings" (Shiller, 1981), and, more problematically, they are horizon dependent. For instance, firms with different characteritics (e.g., value versus growth) likely have different cashflow duration. Earnings also impose a low frequency (quarterly at most) on these regressions, which thus speak to long-term, low-frequency information—they are not designed to detect rapid changes in the information flow that presumably occur at higher frequencies.

We propose a measure that is separate from *1.* earnings data and *2.* the regression approach. Our starting point is to proxy for the "noise" element of the price signal instead (e.g., Dessaint, Foucault, Frésard, and Matray (2018); Honkanen and Schmidt (2021)). Noise is often modeled as supply shocks, reflecting price changes unrelated to information and without which investors would refuse to trade (Milgrom and Stokey, 1982). Following a large literature (Coval and Stafford, 2007; Edmans, Goldstein, and Jiang, 2012), except for a few adjustments (Wardlaw, 2020), we measure supply shocks from mutual fund outflows (MFFlow), attempting to map mutual funds that are subject to large redemptions into unanticipated liquidation "fire sales" of individual stocks they hold. This proxy for "nonfundamental" shocks is horizon independent and can be constructed monthly, thus addressing some of the issues above, but is no panacea either. We do not know whether a fund liquidates a stock in response to the outflow or in response to adverse fundamental information—value and noise remain separately unobservable. We address the incomplete observability of supply shocks by moving away from a regression approach.

The idea is to use the MFFlow measure to create an empirical framework in which the informational content of prices can be cleanly recovered in response to a specific event (presumed fire sales). To be clear, we are not interested in using MFFlow as an instrumental variable in a regression, but only to select a subset of firms in its lowest decile, those subject to the largest supply shocks in a given month (278 firms on average). Each month we form an equally-weighted portfolio of these shocked firms and track their cumulative abnormal returns (CAR) relative to the Carhart (1997) factors and five-industry portfolios following the shock. In response to the supply shock CAR on this portfolio experiences strong downward pressures. Yet, because supply shocks are incompletely observable, investors cannot immediately tell whether fundamental or supply drives the CAR drop they observe. This situation gives a particularly important role to learning from prices: over time it would become increasingly unlikely that the CAR drop was caused by fundamental news and CAR would recover. The path of CARs over this recovery period, which we take to be yearly as most of the reversal from the shock occurs over this period in our sample, is the focus of this paper.

In his presidential address, Darrell Duffie discusses exactly this idea and concludes "*a suitable specialization of the neoclassical model of He and Wang (1995) could be used to analyze price dynamics in this setting*" (Duffie, 2010). This model is indeed easily interpreted in terms of the empirical framework above. The initial date represents the month-end date over which the supply shock takes place and the economy's finite horizon the end of the recovery window over which CARs converge back to fundamentals on average. Investors learn from prices along with their own private information which of value or noise drives the price realizations they see. We specialize this model in two ways that play a central role in the analysis. First, we consider a general process of private information collection, which the average number of private signals across investors entirely subsumes in equilibrium. This average measures the flow of private information and is a deterministic and nondecreasing function of time but is otherwise *arbitrary*. Second, trading takes place continuously over the recovery window: in equilibrium prices follow a diffusion process, with the important consequence that the quadratic variation of their paths is observable.

This theoretical framework delivers a key insight for price informativeness. Equilibrium imposes a precise relation between price informativeness and the flow of private information: when noise is fully transitory the speed at which information flows from prices is proportional to the square of the speed at which private information accumulates. Therefore, the task of recovering the flow of price informativeness is one of recovering the flow of private information: a deterministic but arbitrary function of time and three parameters (the initial level of this function, that of uncertainty regarding fundamentals, and the extent of variation in noise). What is now needed is an identification procedure to recover these model primitives.

The main theoretical result of the paper is a recovery formula for the private information flow. We exploit a well-known property of high sampling frequencies, namely that in continuous time the quadratic variation (QV) of diffusions is observable. In this CARA-normal model equilibrium prices follow a Gaussian process: their diffusion is deterministic, and just a function of the private information flow and the parameters above. Hence, sampling the paths of CARs sufficiently frequently following the shock we could compute their QV from the data, invert this function and recover the private information flow as a function of observed QV, similar to the typical exercise of inverting the Black-Scholes formula to recover implied volatility as a function of options prices. We construct the QV of the portfolio of shocked firms using intraday NYSE TAQ data. Each day over the yearly recovery period we sample portfolio CARs at the 10-minutes frequency, and calculate their QV following Aït-Sahalia, Kalnina, and Xiu (2020). For each shock we obtain a daily time series of QV, which we plug in the recovery formula, which in turn gives a curve of private information flow sampled daily over the yearly window, given the three remaining parameters.

The last step is to identify these three parameters using these daily time series, along with the time series of CARs. In this model the empiricist, who is part of the equilibrium construction, is defined as an external observer who only observes the history of CARs. The usual exercise is to let the empiricist predict each day over the recovery window what the next CAR datapoint will be. In the model she does so by means of Kalman filtering, and infers fundamentals and supply from the paths of CARs following each supply shock. Remaining parameters can then be identified by maximum likelihood estimation. After recovering these parameters we reconstruct the curve of price informativeness from its equilibrium relation with the recovered curve of private information flow, the key insight above. This procedure produces as many such curves for price informativeness sampled daily as there are shocks over their associated yearly recovery period.

The novelty of this procedure is to recover the "shape" of learning from prices, as opposed to a single "snapshot" of it. To summarize this shape with just a few numbers we follow the literature that models the term structure of interest rates and define similarly measures of Level, Slope and Curvature of price informativeness curves. Level measures the empiricist's prior precision regarding fundamentals, and thus its inverse prior uncertainty, which represents the largest amount of information that the empiricist can learn about fundamentals for a given shock. Slope captures the increase in price informativeness over the recovery window, and thus how much information she effectively learns from prices over this period. Curvature quantifies how fast this information becomes available to her, with a concave (convex) curve indicating information flows from prices early (late). We apply this procedure over the last two decades, from 1997 to 2022, a total of 25 years $\times$ 12 months = 300 shocks. The resulting 300 datapoints for each of Level, Slope and Curvature allow us to understand how the shape of price informativeness has evolved over time.

Quite clearly, learning from prices is substantially more difficult during crises. Three crises occur over the sample period: the Dotcom bubble, the global financial crisis, and the Covid-19 pandemic. Level and slope fell sharply on these occasions, with a particularly impressive drop upon the global financial crisis. Not only is there a larger amount of information to be learnt, but less information can be learnt from prices. Curvature also spikes (the information flow becomes strongly convex), meaning that fundamental information is slower to make its way through prices. Interestingly, towards the end of crises the flow of price informativeness reverses equally fast, with a sudden rebound in level and slope along with a sharp drop in curvature: uncertainty declines and the information flow accelerates markedly. This reversal in the shape of price informativeness sounds the death knell of a crisis.

Two trends are unmistakable. Over the sample period, but mostly over the last decade, level rose and slope declined. Whereas the decline in prior fundamental uncertainty (rising level) has been quite steady over the sample period (apart from crises) slope rose during the early part of the sample but started declining distinctly after the financial crisis. Curvature also exhibits a weak upward trend, with curvature swings becoming weaker over time. The picture is, over the last decade prior uncertainty has declined, and prices have not only revealed less information but have also incorporated it more slowly. These trends cannot be explained by changes in characteristics of shocked firms we select, e.g., these becoming smaller, less liquid, less covered by analysts or exhibiting greater price inelasticities (Koijen and Yogo, 2019). In fact, not only do trends persist when controlling for these characteristics but they become statistically stronger—uncertainty has dropped but prices have revealed less information and at a slowler pace.

However, which patterns in price informativeness should we expect in the first place? We examine this question from the perspective of classical views on market efficiency (Shiller, 1981; Black, 1986), which allow us to appreciate what is a low or high level of price informativeness. In his presidential address Fischer Black conjectures "almost all markets are efficient", meaning "price is within a factor 2 of value" at least 90% of the time. This claim can be formulated mathematically in the model, and implies a lower bound on pricing accuracy relative to fundamental prior precision. Unsurpringly, crises are systematically associated with inefficient markets. Yet, what stands out is the long period of substantial inefficiency surrounding the Covid-19 pandemic. Intuition would perhaps suggest prices are more efficient when fundamental precision is higher, but what matters is how it rises relative to pricing accuracy. The rise in level and the decline in slope do not translate in an equivalent improvement in pricing accuracy, meaning these trends in fact result from inefficient prices. Even adopting a different, stronger notion of efficiency (Fama, 1970; LeRoy, 1989), if we accept the idea that prices do not reflect information instantly but appropriately fast, prices have incorporated information significantly more slowly over the last two decades.

Although firm characteristics do not determine aggregate trends, they do play an important role in the shape of price informativeness *across firms*. To a large extent, size and liquidity (orthogonalized to size) are the two characteristics that matter most. Because larger firms exhibit lower fundamental uncertainty (higher level) they offer less potential for learning and less information to be learnt from their price (lower slope). Liquidity exhibits exact opposite patterns relative to size: higher liquidity is associated with a larger and faster amount of information. More liquid firms are also associated with greater fundamental uncertainty to the extent that they are also subject to tremendously less noise trading. Most interestingly, the shape of price informativeness has become increasingly insensitive to size and liquidity. This result may speak to the "data feedback loop" (e.g., Begenau, Farboodi, and Veldkamp (2018) or Veldkamp (2023)), which postulates that large firms benefit more from data, generate more data and thus grow even larger. *Among shocked firms, at a yearly horizon and over the last decade* this mechanism may have become less prevalent.

We position these conclusions and the procedure on which they rely relative to Ebit-based regressions first. Our procedure exploits supply shocks with the benefit of resolving horizon dependence but at the cost of focusing on a subset of shocked firms. Yet, although Ebit-based R^2—how much cross-sectional variation (Bai et al., 2016; Farboodi, Matray, Veldkamp, and Venkateswaran, 2022) or time-series variation (Davila and Parlatore, 2023) in returns is explained by variation in

Ebit—speaks to the substantially broader CRSP universe, it behaves similarly when restricted to our sample of shocked firms. In addition, our procedure uses intraday data with the benefit of recovering curves of price informativeness. Yet, provided we focus on time-series variation (Davila and Parlatore, 2023), R^2 can be mapped into these curves: R^2 summarizes a curve of price informativeness with the ratio of its slope to its terminal point (level plus slope), meaning what is learnt from prices relative to what can be possibly learnt. Ebit-based regressions are intended to capture low-frequency, long-term information and, unsurprisingly, their R^2 is remarkably stable relative to ours, e.g., variation during crises is weak. Magnitudes also differ, but such differences are commonplace in different asset-pricing contexts (e.g., Roll (1988), Morck, Yeung, and Yu (2000) or Cochrane (2011)). However, the two measures are mostly consistent across key firm characteristics.

The microstructure literature offers measures that are designed to pick up high-frequency information, the kind of information "where knowing a moment before others know something is valuable" (Farboodi et al., 2022), e.g., the PIN measure of Easley, Kiefer, O'Hara, and Paperman (1996) and the price jump ratio developed by Weller (2018). In particular, the measure of Weller (2018) is meant to detect the extent of information prices incorporate just a month ahead of an announcement, and PIN is typically computed daily. Although our procedure uses intraday data and samples price informativeness daily, the relevance of the information it captures persists for as long as CARs eventually recover, which occurs at a yearly horizon on average. We think this information frequency puts us between microstructure-based measures and Ebit-based regressions. Our focus on supply shocks is also related to Brogaard, Nguyen, Putnins, and Wu (2022), but the procedure differs, with theirs relying on statistical decompositions and ours on structural estimation.

Section 2 presents the empirical strategy and the construction of the path of CARs following supply shocks. Section 3 interprets these paths in terms of the model and derives the main insight into the equilibrium relation between price informativeness and the private information flow. Section 4 contains the main theoretical result, a recovery procedure for the shape of price informativeness and how to summarize it. Section 5 examines this shape over time, across firm charactersitics, from the perspecitive of market efficiency, and relative to Ebit-based regressions.

2 Price recovery from supply shocks

We present the empirical strategy, how we implement it using standard measures based on mutual fund outflows (supply shocks), and how we "strip out" common factors to obtain the recovery path of cumulative abnormal returns (CARs) following these supply shocks.

2.1 Empirical strategy

The rational-expectations literature (e.g., Hellwig (1980), Diamond and Verrecchia (1981) or Grossman and Stiglitz (1980)) and the microstructure literature (e.g., Kyle (1985) or Glosten and Milgrom (1985)) commonly view prices as signals of the form:

"fundamental value plus noise."

Prices change to reflect new information about payoffs, but they also change for non-informational reasons. Without such "noise" investors would refuse to trade (Milgrom and Stokey, 1982).

A conceptual issue is that noise and value are not observable separately, which regression-based approaches in the literature address by choosing a proxy for the "value" element of the price signal (e.g., Bai et al. (2016), Farboodi et al. (2022) or Davila and Parlatore (2023)). Constructing proxies for value in turn raises practical issues: *1.* because dividends are not by paid by all CRSP firms, earnings, often Ebit at a $1-$ up to $3-$years horizon, are used in lieu of value, and there exists as many such proxies as there are definitions of earnings (e.g., Ebit, free cashflow or net income). *2.* Earnings are horizon dependent, and selecting a single horizon is difficult as firms with different characteristics have different cashflow duration. For instance, a young company that has developed a promising technology will likely see its stock price rise, but this informational increase may not be reflected in Ebit over the subsequent 3 years, as associated profits will only arise later. *3.* Earnings impose a quarterly sampling frequency when regressed on prices, with these regressions thus measuring low-frequency information. Yet, even if information concerns long-term horizons rapid changes in its flow presumably occur at higher frequencies.

Our starting point is to move away from earnings data, and instead construct a proxy for the "noise" element in the price signal. Noise is often modeled as supply shocks and reflects, more broadly, price changes unrelated to information.[1] We measure supply shocks from mutual fund outflows following a large literature that attempts to map mutual funds subject to large redemptions into unanticipated liquidation "fire sales" of individual stocks they hold (Coval and Stafford, 2007; Edmans et al., 2012). This proxy is horizon independent and can be constructed monthly (e.g., Honkanen and Schmidt (2021)), resolving some of the practical issues above. Yet this proxy is noisy, as it remains separately unobservable from value, and is not contemporaneously observable, as redemptions are not disclosed for some time. In particular, we do not know whether a fund manager liquidates a stock in response to the outflow, or because she had adverse fundamental information, or for some other reason, e.g., she liquidates this stock because it is more liquid.

To address the incomplete observability of supply shocks we move away from the regression approach. We use the proxy for supply shocks to create an empirical framework in which the informational content of prices can be recovered in response to the presumed fire sales event. Formally, the only use we make of the proxy is to select a subset of firms that are subject to the largest supply shocks in a given month, those that fall in its lowest decile (278 firms on average). Each month we form an equally-weighted portfolio of these shocked firms and track their cumulative abnormal returns (CAR) relative to the traditional factors. CAR on this portfolio experiences substantial downward pressures upon the shock. Yet, because supply shocks are incompletely

[1]There are different ways to model uninformative price changes: liquidity needs may arise from private investment opportunities (Wang, 1994), investor specific endowment shocks, or income shocks (Farboodi and Veldkamp, 2017), noise trading (Black, 1986), misallocation or market inefficiencies (Grossman, 1995).

observable, investors cannot immediately tell whether the price drop is fundamental- or supply-driven and must learn from prices following the shock: over time it would become increasingly unlikely that the CAR drop was caused by fundamental news and CAR would recover.[2] We explain the construction of CAR paths over this recovery period and the measurement of supply shocks in separate subsections next.

For the purpose of constructing the measure of price pressure caused by fund outflows we collect data from two sources. We obtain data on fund flows from CRSP and s12 holdings data from Refinitiv (formerly CDA Spectrum). Data on stock prices and volume is collected from CRSP and balance-sheet data from Compustat. We start from a stock universe that is made of all US common stocks traded on NYSE, Nasdaq, or AMEX and then select a subset of shocked firms in a way we describe below.[3] The sample period runs from 1997 through 2022.

2.2 Measurement of supply shocks

We start from the methodology of Edmans et al. (2012) (*henceforth*, EGJ), and compute downward price pressure on stock i caused by mutual fund trading in month t as:[4]

$$MFFlow_{i,t} = \sum_{j=1}^{m} \frac{SHARES_{i,j,q(t)-1} \times PRC_{i,t-1}}{TotalAssets_{j,t-1}} \times \frac{Flow_{j,t}}{VOL_{i,t}}. \quad (1)$$

The first ratio is the fraction of wealth fund j holds in stock i at the beginning of month t. The number of $SHARES$ of stock i that fund j holds is recorded at the end of quarter $q(t) - 1$ prior to month t, and we assume these holdings remain constant over each quarter. Price, $PRC_{i,t-1}$, and total net assets, $TotalAssets_{j,t-1}$, are recorded at the end of the preceding month $t - 1$. This position (in % of wealth) is then multiplied by total dollar outflow, $Flow_{j,t}$, fund j experienced in month t, which is negative by construction.[5] The resulting product represents "hypothetical

[2]This is a key difference between mutual fund outflows and, for instance, additions/deletions of stocks in an index. Whereas these additions/deletions also represent nonfundamental demand/supply shocks to stocks (in the sense that index funds are forced to buy/sell these stocks), these changes in index constituents are publicy announced, thus removing the possibility that associated price changes are driven by fundamentals.

[3]That is, all stocks in the CRSP universe with sharecodes (shrcd) 10 or 11, and exchange code (exchcd) of 1, 2, 3, 31, 32, or 33. We additionally exclude firms that have multiple share classes, i.e., non-blank entries in "shrcl" in the CRSP dataset.

[4]EGJ calculate $MFFlow$ at a quarterly frequency, because fund holdings are reported per quarter. Since it is assumed that portfolio holdings of funds remain constant over each quarter, we can compute monthly $MFFlow$ as fund outflows are available at a monthly frequency. This methodology is similar to Honkanen and Schmidt (2021).

[5]Following previous studies using trading pressure induced by flows (e.g., Edmans et al. (2012); Ali, Wei, and Zhou (2011) Hau and Lai (2013)), our focus is on fire sales rather than "fire purchases." This choice is customary, for two reasons. First, mutual funds are more likely to face significant pressure to divest stock positions when outflows deplete their cash reserves. In contrast, they have more discretion in timing stock purchases after receiving inflows. Second, fund inflows, unlike outflows, are known to strongly respond to past fund performance, indicating a convex fund flow-performance sensitivity (see, for instance, Sirri and Tufano (1998) or Gruber (1996)). Therefore, "fire purchases" are hardly nonfundamental nor exogenous.

sales", which assumes that fund j responds to its total outflow by liquidating proportionally to her portfolio at the end of the quarter $q(t) - 1$ prior to month t. This approach is in contrast to that of Coval and Stafford (2007) who use actual sale transactions of mutual funds. A common concern raised against using direct transactions is that distressed funds may choose to sell specific stocks for which they possess unfavorable information (Huang, Ringgenberg, and Zhang, 2023), which could violate the assumption the shock is exogenous and nonfundamental. The use of hypothetical sales serves to mitigate this concern. Furthermore, because there is significant heterogeneity in trading volume across firms, the measure is scaled by $VOL_{i,t}$, which is meant to capture trading volume in dollar units and is calculated by multiplying volume in shares with month-end stock price, $PRC_{i,t}$. Finally, the measure is aggregated across all funds for which percentage monthly outflow is $|Flow_{j,t}/TotalAssets_{j,t-1}| \geq 2.5\%$ (i.e., funds subject to big redemptions).[6]

We further adjust the measure to address a recent critique that *MFFlow* is mechanically contaminated by fundamental information (Wardlaw, 2020). The issue is in the definition of dollar volume *VOL*, which converts volume in shares to dollar units using month-end prices. As a result, the stock price appears both in the numerator and denominator of (1) but at different dates in a way that inverse gross return, $\frac{PRC_{i,t-1}}{PRC_{i,t}}$, is a component of the measure. Thus the measure is not exempt of fundamental information. We address this problem by redefining VOL using *daily* dollar volume, which is now directly available from CRSP, and then cumulate it over the month. This adjustment mitigates Wardlaw (2020)'s criticism as it converts share volume into dollar volume on a daily basis, as opposed to a single month-end conversion.

In Figure 1 we illustrate how this adjustment affects average daily cumulative return, in excess of the CRSP equally-weighted index. In particular, this figure compares average cumulative excess returns of stocks that fall in the lowest decile of the full sample *MFFlow* distribution, under its original definition (red curve) and under our proposed adjustment (blue curve). The red curve reveals impressive downward price pressure on these stocks, with a whooping drop of more than 3% on average in the shocked month (shaded area). Returns then slowly recover back to fundamentals within roughly 1.5 years after the shock. Under our adjusted *MFFlow* the pattern remains similar but magnitudes are weaker, suggesting that Wardlaw (2020)'s critique does matter.

2.3 Recovery path of portfolios of shocked firms

Our focus is quite different from the way the *MFFlow* measure is often used in the literature. The literature focuses on the month (or quarter) the supply shock takes place and commonly uses

[6] EGJ use a threshold of 5% to define "large quarterly outflow." Intuitively, the threshold should become looser as the time period over which outflows are calculated decreases. Therefore, similar to Honkanen and Schmidt (2021), who use a threshold of 2%, we consider monthly outflows of $\geq 2.5\%$ as "large outflows." However, results remain similar for thresholds between 2% and 5%.

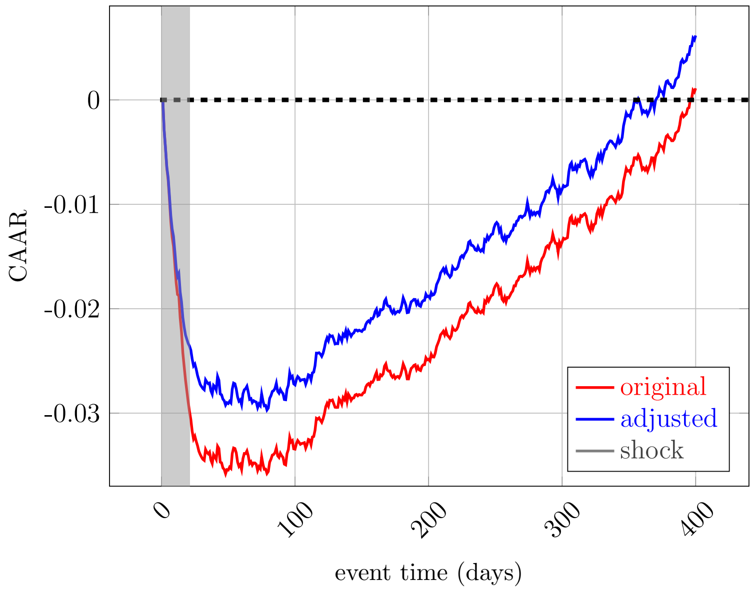

Figure 1: Average cumulative abnormal log returns (CAARs) around supply shocks. This figure illustrates the daily cumulative average log returns, adjusted for the CRSP equally-weighted index, of stocks belonging to the lowest decile of the full sample $MFFlow$ distribution. The shaded area highlights the month in which the shock occurs, and the $x-$axis correspond to days relative to this shock. The curve "original" is based on the $MFFlow$ measure of Edmans et al. (2012), whereas the blue curve is based on the adjusted $MFFlow$ measure we propose to address the criticism in Wardlaw (2020).

MFFlow as an instrumental variable in a regression specification.[7] In contrast, we are exclusively interested in cumulative returns over the recovery period following the month in which the shock occurred, as opposed to *MFFlow* itself: we only use $MFFlow$ to form portfolios of firms.[8] Each month we select stocks that fall in the lowest decile of the monthly MFFlow distribution (because $MFFlow$ is negative the lower it is, the stronger the associated supply shock is). We then form equally-weighted portfolios of these stocks each month, which we hold over a "recovery period" to be described. Finally, each month we record cumulative returns on the portfolio over the recovery window. The portfolio contains 278 stocks on average and, because it is equally weighted, can be interpreted as the "average shocked stock."

We want to isolate the part of portfolio returns that is specific to recovery from the shock, controlling for aggregate variation (time specific) and variation that is stock specific. For instance, we know firms mutual funds hold tend to be smaller stocks (Berger, 2023; Wardlaw, 2020). They

[7]For instance, related literature in empirical corporate finance shows that supply shock impact shareholder activism (Derrien, Kecskes, and Thesmar, 2013; Norli, Ostergaard, and Schindele, 2014), corporate disclosures (Zuo, 2016), takeover attemps (Edmans et al., 2012), corporate investments (Dessaint et al., 2018), option grant timing (Ali et al., 2011) and the use of credit lines (Acharya, Almeida, Ippolito, and Perez, 2014).

[8]In this respect, this paper is closer to Honkanen and Schmidt (2021) who discuss the effect of supply shocks on cross-asset learning from prices in a rational-expectations model or to Lee and So (2017) who show that analyst coverage of stocks increases after supply shocks.

also tend to outperform the equally-weighted CRSP index (Wardlaw, 2020), which is usually used as the main benchmark (see, e.g., Edmans et al. (2012) or Dessaint et al. (2018)). Therefore, it is important to control for dimensions along which shocked firms in the portfolios we form are different from other CRSP firms. Specifically, we compute the cumulative abnormal return (CAR) on the portfolio of shocked firms in excess of its exposure to the usual four factors (Carhart, 1997). In addition, we control for sector-specific effects relative to the five industry portfolios, which along with the market constitute the main determinants of volatility.[9] We estimate factor exposures each day over the recovery window using intraday variation according to the procedure we describe in Section 4, at which point we will clarify the relevance of intraday variation in this computation.

Finally, to select the length of the recovery period we examine the average time it takes for CARs to recover. Intuitively, we expect the recovery period to end once CARs level off. Yet, whereas Figure 1 suggests CARs fully recover after two years, they in fact continue to rise above their pre-shock value when extending the window beyond two years. This continuation likely represents a premium unrelated to the shock itself, but due to certain characteristics of shocked stocks (Wardlaw, 2020) that the factors described above are precisely aimed to capture. In addition, the tendency of fire sales to cluster (firms that get shocked repeatedly) may also contribute to extend the duration of recovery (Honkanen and Schmidt, 2021). To control for fire sales clusters, we run panel regressions, with the additional benefit of controlling for (unobserved) characteristics that do not vary within groups of observations and that the factors above would not subsume. We regress monthly excess returns on a set of event-time dummies, D_{t+x}, for each month t in a window of $x \in [-3, 30]$ months around the shock, including stock- and time-fixed effects and factors listed above (market, size, value, momentum and industry).[10] The coefficients on these dummies capture the marginal average abnormal return on shocked firms in the associated month relative to unshocked firms, which we plot in Figure 2. Abnormal returns recover rapidly first and then *truly* level off. We select the length of the recovery window to be $1-$year, since most of the reversal from the shock occurs over this period. In Section 4 intraday data will confirm that most variation following the shock occurs over one year.

[9]Industry categories are manufacaturing, hitech, consumer health and other. Classifications are based on four-digit SIC codes which we obtain from Kenneth French's data library that is accessible here.

[10]Here we omit the category "other" to prevent perfect collinearity.

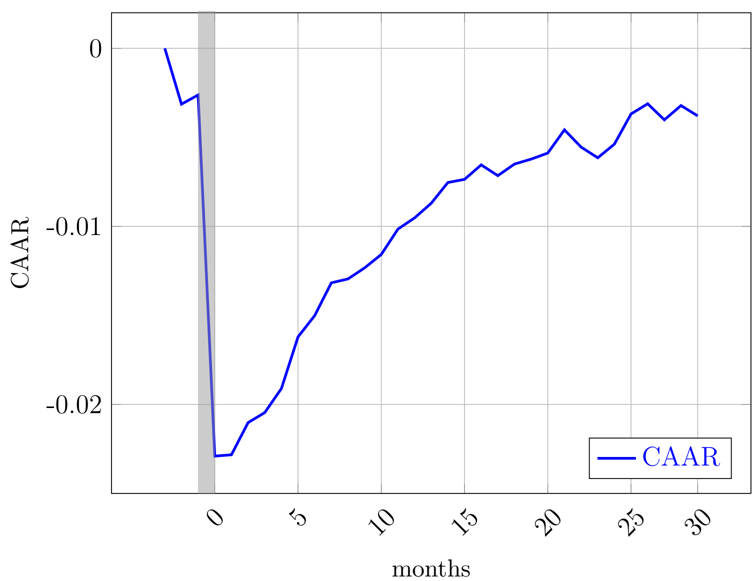

Figure 2: CAARs based on panel regression approach. This figure shows cumulative average abnormal returns (CAARs) of shocked firms in event-time, where date zero is the month over which the fire sale takes place. The blue curve plots the cumulated coefficient estimates on the event-time dummy variables.

3 Model

In his presidential address, Darrell Duffie discusses the possibility that the incomplete observability of supply shocks explains the slow recovery of CAR patterns in Section 2: "*If investors are unable to immediately identify whether the price drop is due to the arrival of adverse information or due to an unanticipated supply shock, then it would take time for them to improve their inference by observing cash flows and other information [...] Over time, the conditional probability that the price drop was due to the arrival of adverse private information would decline and the price would recover in expectation [...] a suitable specialization of the neoclassical model of He and Wang (1995) could be used to analyze price dynamics in this setting.*" (Duffie, 2010) The goal of this section is to provide this specialization. We use a continuous-time extension of He and Wang (1995) in which we keep the process of private information collection as general as possible, along the lines of Cujean (2020). The key insight of this model is that in equilibrium the speed at which information flows from prices is proportional to the square of the speed at which private information accumulates.

3.1 Rational-expectations framework

Time is continuous and runs up to a finite horizon T, at which some unobservable "fundamental" value, $\widetilde{F} \sim \mathscr{N}(0, \tau_F^{-1})$, will be paid. The economy is populated with a continuum of agents indexed by $i \in [0, 1]$, all of whom exhibit CARA utility over terminal wealth with common absolute risk aversion, γ. Trading takes place continuously over $[0, T)$. The market consists of two assets. The first asset is a risky stock with equilibrium price P_t at time t. The stock is a claim to the fundamental $\widetilde{F}$. The second asset is a riskless claim with perfectly elastic supply and a rate of return normalized to $r = 0$ (consumption takes place only once).

The problem of agent i is to find a predictable portfolio strategy θ^i maximizing her expected utility over terminal wealth:

$$\mathbb{E}\left[-e^{-\gamma W_T^i} \,\middle|\, \mathscr{F}_t^i \right] \tag{2}$$

$$\text{subject to} \quad W_T^i = W_0^i + \int_0^T \theta_t^i \mathrm{d}P_t + \theta_T^i \Delta P_T, \tag{3}$$

where $\mathscr{F}_t^i$ denotes agent i's information set at time t to be described. The budget constraint in Eq. (2) includes a price discontinuity of size ΔP_T occurring at the horizon date (see Cujean (2020)).

Prices change to reflect informed trades but also due to unobservable shocks to the supply of the stock, $\widetilde{m}$, which evolves according to an AR(1) process:

$$\mathrm{d}\widetilde{m}_t = -b_m \widetilde{m}_t \mathrm{d}t + \tau_m^{-1/2} \mathrm{d}B_{m,t}, \quad \widetilde{m}_0 \sim \mathscr{N}(0, \infty) \tag{4}$$

with B_m a Brownian motion, τ_m and b_m constants, which represent supply precision and mean reversion in supply, respectively. This is the usual noise trading story ($\widetilde{m}$ is the supply available to the market, whereas noise traders—agents who trade for reasons unrelated to fundamental information—have inelastic demands of $1-\widetilde{m}$ units of the stock (in total supply of 1)); accordingly, the coefficient b_m can be viewed as the fraction of the demand of noise traders who revert their trades over a time interval $\mathrm{d}t$. Thus, $b_m \equiv 0$ means that, unlike informed traders, noise traders are buy-and-hold investors. For instance, this assumption is appropriate for the purpose of explaining intraday patterns. Finally, we assume diffuse priors regarding the initial supply shock.

This model is easily interpreted in terms of the empirical framework of Section 2. Date 0 represents the month-end date over which the initial supply shock, $\widetilde{m}_0$, takes place and $[0,T]$ is the recovery window over which prices converge back to fundamentals $\widetilde{F}$ (1 year in expectation as illustrated in Section 2). Investors cannot immediately tell which of fundamentals, $\widetilde{F}$, or initial supply shock, $\widetilde{m}_0$, drives the first price realization, P_0. Over time they learn from subsequent prices along with "other information" they observe to decide which of the two was responsible for P_0. We now specify how this "other information" is collected.

3.2 Private information collection

The goal is to maintain a flexible structure for investors' private information. We assume that over time each agent i obtains an increasing sequence of private signals about the fundamental:

$$S^i_j = \widetilde{F} + \epsilon^i_j, \quad j = 1, ..., n^i_t \tag{5}$$

where $n^i_t \in \mathbb{N}^*$ denotes the number of signals agent i has collected up to time t and where $\epsilon^i_j \sim \mathscr{N}(0, \tau_S^{-1})$ represents the "idiosyncratic" noise in agent i's $j-$th signal. By *idiosyncratic* we mean there is one such random variable per agent i and signal j, and that these random variables are sufficiently independent for a version of the Strong Law of Large Numbers to hold across agents and signals (e.g., Duffie and Sun (2007)).[11]

Agent i starts with an initial, idiosyncratic number n^i_0 of signals, which is drawn from a distribution π_0 with support $\mathbb{N}^*$. She then collects new signals at arrival times of an idiosyncratic Poisson process $(N^i_t)_{t\geq 0}$ with time-varying intensity $\eta_t(n^i_{t-})$. The intensity at which she gets new signals potentially depends on her current number n^i_{t-} of signal, e.g., an agent who has gathered many signals may be more efficient at collecting new ones in the future.

An agent who collects new signals at time t receives a chunk $(S^i_{j+n^i_{t-}} : 1 \leq j \leq \Delta n^i_t)$, with the incremental number Δn^i_t of signals drawn from a distribution $\pi_t(\cdot; n^i_{t-})$, which potentially depends on her current number n^i_{t-} of signals. Since individual signals are Gaussian and independent (conditional on $\widetilde{F}$), an agent's average new signal:

$$Y^i_t \equiv \tfrac{1}{\Delta n^i_t} \textstyle\sum_{j=1}^{\Delta n^i_t} S^i_{j+n^i_{t-}} = \widetilde{F} + (\tau_S^{1/2}\sqrt{\Delta n^i_t})^{-1}\epsilon^i_t, \quad \epsilon^i_t \sim \mathrm{IID}\mathscr{N}(0,1) \tag{6}$$

[11]That is, for almost every pair (i, i') of agents ϵ^i_j and $\epsilon^{i'}_{j'}$ are pairwise independent for all j and j'.

is a sufficient statistic for the chunk she receives. Thus, an agent's private information is completely summarized by two numbers at any time t, her average signal and her total number n_t^i of signals. In addition each agent observes the history of prices, $(P_t)_{t\geq 0}$, which is endogenous and publicly available and together with their collection of signals accounts entirely for an agent i's information:

$$\mathscr{F}_t^i = \sigma\left(\left(P_s, S_j^i\right) : 0 \leq s \leq t, 1 \leq j \leq n_t^i\right). \tag{7}$$

This process of information collection generates a cross-sectional distribution of number of signals, which we denote by $\mu_t(n)$. This distribution keeps track of the number n of signals across the population of agents at every date t, and satisfies:

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\mu_t(n) = -\eta_t(n)\mu_t(n) + \textstyle\sum_{m=1}^{n-1} \eta_t(n-m)\mu_t(n-m)\pi_t(m; n-m), \quad \mu_0(n) = \pi_0(n). \tag{8}$$

The first term on the right-hand side in Eq. (8) is the rate at which agents leave their type (the fraction of agents of type n who received new signals and thus no longer hold n signals). The second term represents the rate at which agents become of type n.

The key statistic for the equilibrium construction is the cross-sectional average number of signals:

$$\phi_t \equiv \tau_S \cdot \sum_{n\in\mathbb{N}} \mu_t(n)n, \tag{9}$$

the first moment of the cross-sectional distribution. This average represents the flow of private information among agents. To be clear, the flexibility of this information structure is only possible to the extent that the distribution $\pi(\cdot)$ of incremental signals, the intensity $\eta(\cdot)$ at which these signals are gathered and their cross-sectional distribution $\mu(\cdot)$ do not intervene directly in the equilibrium construction as long as they do not depend on the aggregate states, $\widetilde{F}$ and $\widetilde{m}$: the only element that matters is the flow, ϕ, which entirely subsumes these elements and is thus central.

3.3 A key insight for price informativeness in equilibrium

In a noisy rational-expectations model the empiricist is part of the equilibrium construction. The empiricist is someone who only observes *c*ommonly available information:

$$\mathscr{F}_t^c = \sigma(P_s : 0 \leq s \leq t), \tag{10}$$

which in this model is the history of prices. From now onwards we adopt the perspective of the empiricist and view $\mathscr{F}^c$ as containing the time series of CAR we have constructed in Section 2. With learning from prices and rational expectations the empiricist, too, can infer which of fundamental value, $\widetilde{F}$, or initial supply shock, $\widetilde{m}_0$, drove the initial price realization, P_0. At any date t she infers her own estimate, $\widehat{F}_t^c$ and $\widehat{m}_t^c$, of what fundamentals and contemporaneous supply are, respectively.

Most importantly, she computes her estimate of fundamentals with standard error precision:

$$\tau_t^c \equiv \mathbb{V}\left[\widetilde{F}\middle|\mathscr{F}_t^c\right]^{-1}, \tag{11}$$

which defines how informative the history of prices up to time t is, the main focus of this paper.

Theorem 1. *In equilibrium price informativeness at time t satisfies:*

$$\tau_t^c = \tau_F + \frac{\tau_m}{\gamma^2}\int_0^t \left(\frac{\mathrm{d}\phi_u}{\mathrm{d}u} + b_m \cdot \phi_u\right)^2 \mathrm{d}u. \tag{A}$$

The key message is that in equilibrium the extent of persistence, $b_m \geq 0$, in noise trading (the fraction of noise traders who revert their trades over the next time interval $\mathrm{d}t$) determines the extent to which price informativeness accumulates with the square of the slope or the level of the private information flow, ϕ, respectively. Thus, a situation in which Theorem 1 delivers a particularly strong insight is when noise increments are IID ($b_m \equiv 0$, noise is fully transitory), which is likely to happen at high frequency. In this case, the speed at which information flows from prices is proportional to the square of the speed at which private information accumulates.

The case of persistent noise, $b_m > 0$ is also interesting, as it can explain sluggishness in the CAR patterns of Section 2, which only fully recover after approximately 1 year on average. Persistence in supply gives informed investors an incentive to trade on short-term price swings, as opposed to long-term fundamentals, because they know a fraction of noise traders will revert their trades next period (Cespa and Vives, 2012). However, we will focus on the case $b_m \equiv 0$ in which noise is fully transitory because this case delivers simple expressions for equilibrium price coefficients. In particular, in equilibrium the empiricist's estimates of fundamentals and contemporaneous supply satisfy:

$$\widehat{F}_t^c = \frac{1}{\tau_t^c} \cdot \int_0^t (\tau_u^c)' \cdot \frac{\mathrm{d}\tau_u P_u}{\tau_u'}, \tag{12}$$

$$\widehat{m}_t^c = \frac{\tau_t'}{\tau_t} \cdot (\widehat{F}_t^c - P_t), \tag{13}$$

where τ_t denotes investors' average precision at date t:

$$\tau_t \equiv \int_0^1 \mathbb{V}[\widetilde{F}|\mathscr{F}_t^i]^{-1}\mathrm{d}i = \tau_t^c + \phi_t. \tag{14}$$

The rescaled price signal in Eq. (12) takes the intuitive form of "value plus noise", the main insight we started Section 2 with:

$$\frac{\mathrm{d}\tau_t P_t}{\tau_t'} = \widetilde{F} \cdot \mathrm{d}t - \left(\frac{\tau_m}{\gamma} \cdot \phi_t'\right)^{-1/2} \cdot \mathrm{d}B_{m,t}. \tag{15}$$

As the diffusion of the price signal suggests, in a noisy rational-expectations equilibrium noise precision and risk aversion are not identifiable separately but only as the combination:

$$a \equiv (\tau_m^{1/2}/\gamma)^2, \tag{16}$$

meaning that aggregate variation in the model is measured in units of absolute risk aversion. For later convenience (parameter estimation in the next section) we introduce the following parameter transformations, which take noise in units of absolute risk aversion, $1/a$, as a *numéraire*:

$$b \equiv a \cdot \tau_F \quad \text{and} \quad c \equiv b + a \cdot \phi_0 \ . \tag{17}$$

Therefore, from Theorem 1, the task of recovering the flow of price informativeness is one of recovering the flow of private information, $(\phi_t)_{t>0}$, and the three parameters:

$$\boldsymbol{\Theta} \equiv \begin{pmatrix} a & b & c \end{pmatrix}. \tag{18}$$

4 Recovering the shape of learning from prices

The novelty of the method we propose is to recover the shape of learning from prices, as opposed to a single "snapshot" of it. This is possible for two reasons. First, the empirical framework we adopt identifies an event (presumed fire sales) from which prices (CARs) recover along the lines of the nonstationary model of Section 3. Second, sampling the paths of CARs at high frequency (intraday) and calculating their associated quadratic variation allows us to recover the flow of private information up to the parameter $\boldsymbol{\Theta}$ from the quadratic variation implied by the model, similar to the typical exercise of inverting the Black-Scholes formula to recover implied volatility. We identify the remaining parameter $\boldsymbol{\Theta}$ by maximum likelihood (ML), letting the empiricist infer fundamentals and supply from the paths of CARs over the recovery window following each supply shock. Plugging in turn the recovered flow into the formula of Theorem 1 produces as many curves for price informativeness as there are supply shocks. To summarize these curves with just a few numbers we follow the literature that models the term structure of interest rates and define similarly measures of *Level*, *Slope* and *Curvature* of the shape of price informativeness.

4.1 Recovering private information flow from quadratic variation

We obtain a model-based "recovery theorem" for the model primitive function ϕ (the private information flow) using a property of high-sampling frequencies. We start by writing the evolution of equilibrium prices under the empiricist's view:

$$\mathrm{d}P_t = \tau_t'/\tau_t \cdot (\widehat{F}_t^c - P_t) \cdot \mathrm{d}t - a^{-1/2} \cdot \tau_t^{-1} \cdot \left(1 + a \cdot \phi_t'\right) \cdot \mathrm{d}\widehat{B}_t^c, \tag{19}$$

where $\widehat{B}^c$ is a Brownian motion under the empiricist's probability measure. From the above we introduce the following definition.

Definition 1. *The diffusion of CAR in equilibrium at date t satisfies:*

$$\underbrace{\sigma_t \equiv \sqrt{\mathrm{d}\langle P_t\rangle/\mathrm{d}t}}_{\textit{data-implied } \sqrt{QV}} = \underbrace{a^{-1/2} \cdot \tau_t^{-1} \cdot \left(1 + a \cdot \phi_t'\right)}_{\textit{model-implied } \sqrt{QV}}. \tag{20}$$

In continuous time quadratic variation (QV) of CARs is observable. Hence, sampling the data frequently enough we can recover the function ϕ from QV. Formally, we differentiate the idendity in Definition 1 once and re-arrange it to obtain a second-order ODE for ϕ:

$$0 = a \cdot \sigma_t \cdot \phi_t'' - \left(1 + a \cdot \phi_t'\right)\left(a \cdot \phi_t' \cdot \sigma_t^2 + \sigma_t'\right). \tag{21}$$

The solution to this ODE is analytical and highlighted in the theorem below, the main theoretical result of the paper.

Theorem 2. ***(Recovery)*** *Let σ_t in Definition 1 be a given, observable function of time (the square root of QV). The flow of private information is recovered as:*

$$a \cdot \phi_t(\mathbf{\Theta}) = c - b + \int_0^t \left(\frac{\exp\left(-\int_0^s \sigma_u/a^{1/2}\mathrm{d}u\right)\sigma_s/a^{1/2}}{1/c - \int_0^s \exp\left(-\int_0^v \sigma_u/a^{1/2}\mathrm{d}u\right)\sigma_v^2/a\mathrm{d}v} - 1 \right) \mathrm{d}s. \tag{B}$$

This function is nondecreasing (and finite) if:

$$\left(\inf_{t\in[0,T]} e^{-\int_0^t \frac{\sigma_u}{a^{1/2}}\mathrm{d}u} \frac{\sigma_t}{a^{1/2}} + \int_0^t e^{-\int_0^v \sigma_u/a^{1/2}\mathrm{d}u} \frac{\sigma_v^2}{a}\mathrm{d}v \right)^{-1} \leq c < \left(\int_0^T e^{-\int_0^v \sigma_u/a^{1/2}\mathrm{d}u} \frac{\sigma_v^2}{a}\mathrm{d}v \right)^{-1}. \tag{22}$$

Let $\{a_k\}$ be the set of solutions (if any) to the following equation:

$$\sigma_T = \int_\tau^T e^{\int_v^T \sigma_u/a^{1/2}du} \sigma_v' dv. \tag{23}$$

Then the interval in (22) *is nonempty if $a \geq \max_k a_k$ (and possibly on other regions).*

Theorem 2 shows that the flow of private information, ϕ, is recovered from QV up to the parameter $\mathbf{\Theta}$. In particular, as a result of the parameter transformation in (17) the parameter a now appears systematically multiplying QV, thus acting as a "scaling parameter" on QV. Interestingly, monotonicity on ϕ imposes a lower and an upper bound on $c \in \mathbf{\Theta}$, investors' initial average precision on fundamentals (in units of noise, $1/a$). In the empirical implementation we show these bounds can be remarkably tight (mostly during crises), thus informative as to how high or low this precision can be. Finally, since these bounds depend on $a \in \mathbf{\Theta}$, a parameter to be estimated, the interval in (22) could be empty for certain values of a. Specifically, if (23) has no solution then this interval is never empty, and if instead (23) has (possibly many) solutions $\{a_k\}$ then the interval is nonempty provided a is not "too low" (noise is not "too high") in the sense that $a \geq \max_k a_k$. In the empirical implementation we find (23) always has a unique solution.

For the recovery formula of Theorem 2 to be empirically helpful we need time series of QVs, which requires data at high frequency. We collect intraday transaction prices from the New York Stock Exchange Trade and Quotes (NYSE TAQ) database. Our cleaning procedure follows the approach of Barndorff-Nielsen, Reinhard Hansen, Lunde, and Shephard (2009) and removes entries with transaction prices of zero, corrected trades (indicated by CORR = 0), and entries with abnormal sale conditions (trades for which COND includes a letter code other than 'E' and 'F'). We identify stocks by PERMNO, as opposed to ticker, because CRSP ticker often differ from those in TAQ and because companies frequently change their TAQ tickers (e.g., due to M&A). We merge TAQ tickers with PERMNO using the TAQ CRSP linking table available from WRDS. Intraday prices in NYSE TAQ are not adjusted for dividends or stock splits occurring overnight. To compute adjusted overnight returns, we follow Aït-Sahalia et al. (2020), which ensures daily returns from CRSP and the aggregation of intraday returns from TAQ data are aligned.[12]

To construct the quadratic variation of the portfolio of shocked firms (described in Section 2), we index each day over the recovery period by $t \in [0, T]$. Within each day there are 6.5 hours of trading, and we sample intraday returns n times over this time period. We denote the intraday sampling frequency by $1/n$, which in most applications is taken to be 10-minutes, or $n = 40$ intraday observations (including overnight returns). Approximating P in the model with the log-price of the portfolio we obtain intraday returns as:

$$\Delta_{t,i}^n P \equiv P_{t-1+i/n} - P_{t-1+(i-1)/n}, \quad i = 1, ..., n. \tag{24}$$

To handle asynchronicities in the data we use the previous-tick interpolation method to ensure continuity in the data. Specifically, we choose $P_{t-1+i/n} = P_{t-1+i*/n}$, where $t-1+i*/n$ is the largest observation time before and including time $t-1+i/n$ (Hayashi and Yoshida, 2005). In addition, because in the model prices in equilibrium are diffusion processes—their path exhibits no discontinuities—we remove jumps following one of the two standard techniques in the literature (Mancini, 2001, 2009). Specifically, every day t we compute the following threshold:

$$v_{t,n} = 4\sqrt{\mathrm{BV}_{t,n}}n^{-0.49}, \tag{25}$$

where $\mathrm{BV}_{t,n} \equiv \frac{\pi}{2}\sum_{i=2}^{n} |\Delta_{t,i}^n P||\Delta_{t,i-1}^n P|$ denotes the bipower variation (Barndorff-Nielsen and Shephard, 2004) of log-prices on day t. We then collect the continuous part of the portfolio returns on

[12] We use CRSP (unadjusted) open and close prices to calculate intraday open-to-close returns. Combining these with adjusted (close-to-close) returns from CRSP, we can infer adjusted overnight returns (close-to-open). Since CRSP open and close prices are carefully selected based on additional information beyond just the sequence of trades, they are used as open and close prices for our dataset.

day t in the following vector:

$$\mathbf{Y}_t = \begin{pmatrix} \Delta_{t,1}^n P \cdot \mathbf{1}_{|\Delta_{t,1}^n P| \le v_{t,1}} \\ \vdots \\ \Delta_{t,n}^n P \cdot \mathbf{1}_{|\Delta_{t,n}^n P| \le v_{t,n}} \end{pmatrix}. \tag{26}$$

In Section 2 we are in fact focusing on the portfolio's *abnormal* returns in excess of factors that the literature commonly views as capturing aggregate variation. We proceed analogously intraday and remove aggregate variation from the portfolio's quadratic variation following Aït-Sahalia et al. (2020). Let $\Delta_{t,i}^n X_k$ denote intraday log-returns on factor k and let there be K of them. Repeating the steps above for each individual factor every day we compute a threhold $v_{t,n}^k$ above which absolute returns on factor k are considered discontinuous, and where bipower variation on factor k is computed analogously. We then gather the continuous part of intraday log-returns on day t across factors in a $(n \times K)-$matrix:

$$\mathbf{X}_t = \begin{pmatrix} \Delta_{t,1}^n X_1 \cdot \mathbf{1}_{|\Delta_{t,1}^n X_1| \le v_{t,1}^1} & \dots & \Delta_{t,1}^n X_K \cdot \mathbf{1}_{|\Delta_{t,1}^n X_K| \le v_{t,1}^K} \\ & \vdots & \\ \Delta_{t,n}^n X_1 \cdot \mathbf{1}_{|\Delta_{t,n}^n X_1| \le v_{t,n}^1} & \dots & \Delta_{t,n}^n X_K \cdot \mathbf{1}_{|\Delta_{t,n}^n X_K| \le v_{t,n}^K} \end{pmatrix}. \tag{27}$$

Regressing factors on portfolio returns using intraday variation we obtain on every day t a vector of betas over the recovery period:

$$\widehat{\boldsymbol{\beta}}_t = (\mathbf{X}_t'\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Y}_t, \quad t = 1, ..., T. \tag{28}$$

Finally, we compute a daily time series of the continuous part of residual quadratic variation as:

$$\widehat{\sigma}_t^2 = \mathbf{Y}_t'\mathbf{Y}_t - \widehat{\boldsymbol{\beta}}_t'\mathbf{X}_t'\mathbf{X}_t\widehat{\boldsymbol{\beta}}_t, \quad t = 1, ..., T \tag{29}$$

which we use to proxy for QV, σ_t^2, the main input to the recovery formula of Theorem 2.

As an illustration we plot in Figure 3 the average (across the 300 shocks of the sample) of QV over the recovery window, normalizing by the average level of QV after 200 days following the shock. On average QV of the portfolio of shocked firms is 6% higher immediately after the shock relative to what it will be 200 days later; it declines fast within the next 100 days following the shock, confirming that most variation in returns occurs over a short time period. We keep in mind, however, that the pattern in Figure 3 is only an average, and that individual patterns for each shock may look significantly different.

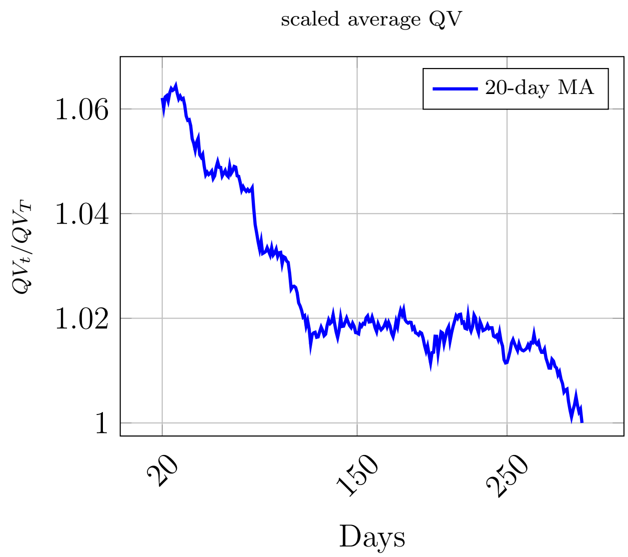

Figure 3: Average QV across recoveries. This plot illustrates the average 20-day moving average of QV, normalized by the level at 200 days after the shock (T=300).

4.2 Identifying remaining model primitives $\mathbf{\Theta}$

At this stage we have constructed time series of portfolio CARs, $\{P_{n\cdot\Delta}\}_{n=0}^{N}$, and of portfolio QV, $\{\sigma_{n\cdot\Delta}\}_{n=0}^{N}$, sampled at a discrete frequency, $\Delta = T/N$. Given the procedure above this frequency is daily, $N = 252$. Thanks to Theorem 2 we recover time series of private information flow, $\{\phi_{n\cdot\Delta}(\mathbf{\Theta})\}_{n=0}^{N}$, and thanks to Theorem 1 we then recover time series of price informativeness, $\{\tau^c_{n\cdot\Delta}(\mathbf{\Theta})\}_{n=0}^{N}$, *for parameter* $\mathbf{\Theta}$ *given*. The remaining task is to identify $\mathbf{\Theta}$ (exactly 3 numbers) using these daily time series.

The usual exercise is to let the empiricist each day $(n-1)\cdot\Delta$, and over the N days that span the recovery window, predict what the next CAR datapoint will be, $\mu_{n\cdot\Delta|(n-1)\cdot\Delta} \equiv \mathbb{E}[P_{n\cdot\Delta}|\mathscr{F}^c_{(n-1)\cdot\Delta}]$, with error variance, $\Sigma_{n\cdot\Delta|(n-1)\cdot\Delta} \equiv \mathrm{Var}[P_{n\cdot\Delta}|\mathscr{F}^c_{(n-1)\cdot\Delta}]$, conditional on the history of CARs, $\mathscr{F}^c_{(n-1)\cdot\Delta}$, up to this day. In this Gaussian framework and in discrete time (daily) the log-likelihood of a given path of CARs, $\{P_{n\cdot\Delta}\}_{n=0}^{N}$, up to time T relative to these predictions is:

$$\mathscr{L}_T \equiv -\frac{1}{2}\sum_{n=1}^{N}\left(\log(2\cdot\pi) + \log(\Sigma_{n\cdot\Delta|(n-1)\cdot\Delta})\right) - \ell(\mathbf{\Theta}), \tag{30}$$

where $\ell(\cdot)$ is the Sum of Squared Errors (SSE) of these predictions defined in the proposition below.

Another benefit of constructing QV at high frequency and mapping it into the diffusion of CARs through the ODE in (21) is that (see Appendix A.3 for a precise statement):

$$\Sigma_{n\cdot\Delta|(n-1)\cdot\Delta} \approx \sigma^2_{n\cdot\Delta}\cdot\Delta, \tag{31}$$

where the approximation is due to the discretization (it holds exactly in continuous time, which is the meaning of Definition 1). Hence, because we observe QV (the right-hand side), maximizing the log-likelihood in (30) is equivalent to minimizing the SSE of the Kalman filter.[13]

Proposition 1. *Given time series of QV, $\{\sigma_{n\cdot\Delta}\}_{n=0}^{N}$, and of CARs, $\{P_{n\cdot\Delta}\}_{n=0}^{N}$, the Sum of Squared Errors (SSE) of the Kalman filter satisfies:*

$$\ell(\boldsymbol{\Theta}) \equiv \frac{1}{2}\sum_{n=1}^{N}\left(\frac{P_{n\cdot\Delta} - \mu_{n\cdot\Delta|(n-1)\cdot\Delta}}{\sigma_{(n-1)\cdot\Delta}\sqrt{\Delta}}\right)^2, \tag{33}$$

where the empiricist's posterior prediction of next CAR datapoint is:

$$\mu_{n\cdot\Delta|(n-1)\cdot\Delta} = \frac{a\cdot\tau_{(n-1)\cdot\Delta}}{a\cdot\tau_{n\cdot\Delta}}P_{(n-1)\cdot\Delta} + \left(1 - \frac{a\cdot\tau_{(n-1)\cdot\Delta}}{a\cdot\tau_{n\cdot\Delta}}\right)\widehat{F}^c_{(n-1)\cdot\Delta}, \tag{34}$$

and where the empiricist's posterior estimate of fundamentals is:

$$\widehat{F}^c_{n\cdot\Delta} = \frac{1}{a\cdot\tau^c_{n\cdot\Delta}}\sum_{k=0}^{n}(a\cdot\tau^c_{k\cdot\Delta} - a\cdot\tau^c_{(k-1)\cdot\Delta})\frac{a\cdot\tau_{k\cdot\Delta}P_{k\cdot\Delta} - a\cdot\tau^c_{(k-1)\cdot\Delta}P_{(k-1)\cdot\Delta}}{a\cdot\tau_{k\cdot\Delta} - a\cdot\tau_{(k-1)\cdot\Delta}}. \tag{35}$$

Empiricist's and average investors' precisions are discretized similarly as:

$$a\cdot\tau^c_{n\cdot\Delta} = b + \sum_{k=1}^{n}(a\cdot\phi_{n\cdot\Delta}(\boldsymbol{\Theta}) - a\cdot\phi_{(n-1)\cdot\Delta}(\boldsymbol{\Theta}))^2/\Delta \tag{36}$$

$$a\cdot\tau_{n\cdot\Delta} = a\cdot\tau^c_{n\cdot\Delta} + a\cdot\phi_{n\cdot\Delta}(\boldsymbol{\Theta}), \tag{37}$$

where $a\cdot\phi_{n\cdot\Delta}(\boldsymbol{\Theta})$ is given in Theorem 2.

Formally, identifying remaining parameters $\boldsymbol{\Theta}$ through Maximum Likelihood (ML) is equivalent

[13]In continuous time the sum of squared errors using integration by parts and removing the constant is:

$$\mathscr{L}_T \equiv \int_0^T \left(\frac{\overline{\tau}'_t/\overline{\tau}_t\cdot(\widehat{F}^c_t - P_t)}{\overline{\sigma}_t}\right)^2 \mathrm{d}t - 2\int_0^T \frac{\overline{\tau}'_t/\overline{\tau}_t\cdot(\widehat{F}^c_t - P_t)}{\overline{\sigma}^2_t}\mathrm{d}P_t. \tag{32}$$

This formulation of the ML problem is standard in this context (e.g., Liptser and Shiryaev (2001)) because quadratic variation is known. In particular, the literature defines a "reference measure" $\mathbb{P}_0$ under which the CAR process does not depend on parameters $\boldsymbol{\Theta}$. This measure $\mathbb{P}_0$ is usually taken to be that under which $\sigma^{-1}\mathrm{d}P$ is a Brownian motion. The log-likelihood function then corresponds to the Radon-Nikodym derivative, $\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{P}_0|_{\mathscr{F}^c_T}$ and the objective in Eq. (32) corresponds to the negative of the log of it.

to solving the following constrained least-squares problem:

$$\min_{\boldsymbol{\Theta}\in\mathbb{R}_+^3} \ell(\boldsymbol{\Theta}) \tag{C}$$

$$\text{s.t. } 1/c \le \min_{n\in\{0,\dots,N\}} e^{-\sum_{k=1}^{n}\sigma_{n\cdot\Delta}/a^{1/2}\cdot\Delta}\sigma_{n\cdot\Delta}/a^{1/2} + \sum_{k=1}^{n} e^{-\sum_{l=1}^{k}\sigma_{l\cdot\Delta}/a^{1/2}\cdot\Delta}\sigma_{k\cdot\Delta}^2/a\Delta \tag{38}$$

$$1/c > \sum_{n=1}^{N} e^{-\sum_{k=1}^{n}\sigma_{k\cdot\Delta}/a^{1/2}\cdot\Delta}\sigma_{n\cdot\Delta}^2/a\cdot\Delta \tag{39}$$

$$a \ge \max_k a_k \text{ solving } \sigma_{N\cdot\Delta} = \sum_{n=1}^{N} e^{\sum_{k=n}^{N}\sigma_k/a^{1/2}\Delta}(\sigma_{k\cdot\Delta}-\sigma_{(k-1)\cdot\Delta}), \tag{40}$$

where the constraints (38) and (39) are discretized versions of (22) and the equation to be solved by $\{a_k\}_k$ in (40) is a discretized version of (23). We solve this problem numerically using a global optimization technique (e.g., `Differential Evolution` in Mathematica) under some regularization.[14] We winsorize QV and abnormal returns at the conventional 1% level.

Program (C) concludes the recovery procedure, and we can now illustrate how it works with a specific shock. Consider the shock that occurred on February 2016 with an associated recovery period running through February 2017, and involving a portfolio of 272 shocked firms. The first column of Figure 4 plots the associated CAR (upper panel) and the square root of quadratic variation constructed in Section 4.1 (lower plot). We see that CARs level off at 10% and QV peaks at 12% following the shock and then drops progressively over the next 100 days. Solving program (C) we recover the parameter $\boldsymbol{\Theta}$, which we plug along with the time series of QV in the recovery formula (B), which gives the red curve for the private information flow (middle panel). We then plug this curve in formula (A), which gives the blue curve for price informativeness. This procedure produces as many such curves for price informativeness as there are shocks (25 years $\times$ 12 months = 300 shocks) over their associated yearly recovery period.

4.3 Summarizing the shape of price informativeness

We would like to describe the 300 curves of price informativeness the procedure delivers (such as the one illustrated in the right panel of Figure 4) in terms of summary statistics. A large literature in fixed income faces a similar task in describing the term structure of interest rates, which is commonly summarized with its level, slope and curvature. We define these three statistics next.

[14]Regularization concerns the upper bound in (22), which holds with strict inequality. Approaching this upper bound will cause the recovered function ϕ_t to explode at T. To prevent this outcome we introduce a quadratic penalty in the SSE objective (C) on approaching this upper bound, setting the penalty parameter in a way that it does not dominate the objective while preventing explosion.

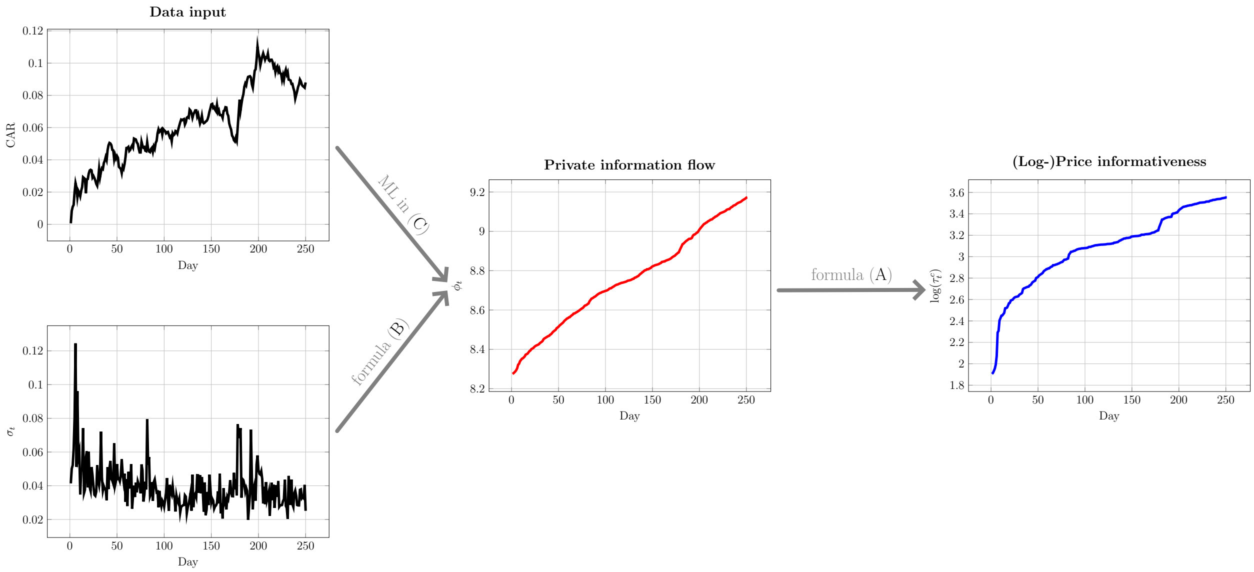

Figure 4: Illustration of methodological steps. This figure illustrates the recovery method for a given shock, which occurred on February 2016. The first column represents data inputs, which consist of CAR (upper panel) and the square root of quadratic variation constructed in Section 4.1 (lower plot). To get to the second column we apply the recovery formula (B) and solve the constrained least-squares problem in (C), which gives the curve for the private information flow (in red). To get to the third column we apply formula (A), which gives the curve for price informativeness (in blue).

Definition 2. *Measuring price informativeness in logs over the recovery window* $(t, t+T]$ *following the shock at month* t*, its level (L), slope (S) and curvature (C) are:*

$$L_t \equiv \log(\tau_t^c) \equiv \log(\tau_F), \tag{41}$$

$$S_t \equiv \log(\tau_{t+T}^c) - \log(\tau_t^c), \tag{42}$$

$$C_t \equiv \log(\tau_{t+T}^c) + \log(\tau_t^c) - 2 \cdot \log(\tau_{t+T/2}^c). \tag{43}$$

In Figure 5 we illustrate what these statistics are meant to capture. Note first that each statistic measures price informativeness in logs, as its magnitude may vary substantially across shocks. The black line represents a curve of price informativeness for a given shock over the yearly recovery window. Its level (in red in the left panel) is the empiricist's prior precision regarding fundamentals (in logs), one of the parameters $\mathbf{\Theta}$ we estimate in Section 4.2. Its inverse, $1/\tau_F$, thus represents the empiricist's prior uncertainty upon shocks, and can be thought of as the *largest amount of information that she can possibly learn about fundamentals* for a given shock. Slope (middle panel) measures the increase in price informativeness over the recovery window, and thus really captures *how much information the empiricist effectively learns from prices* over this period. Finally, curvature (right panel) captures *how fast this information becomes available* to the empiricist. If the curve is concave (convex), information flows from prices early (late), and thus it is easier (more difficult) for the empiricist to obtain quick information.

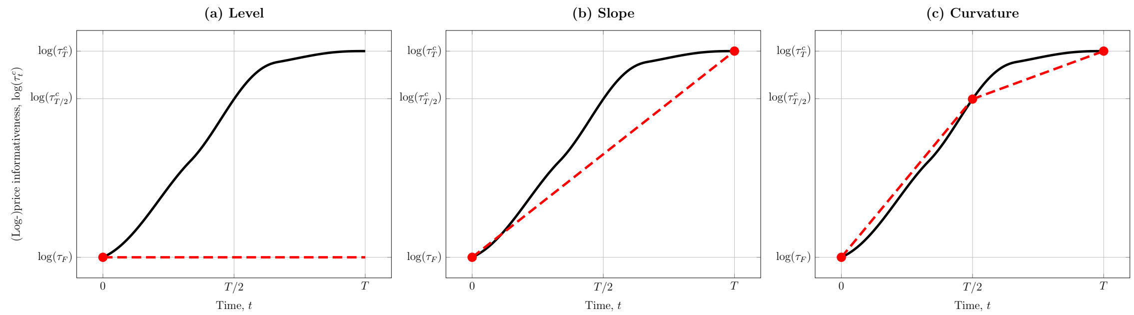

Figure 5: Level, slope and curvature of price informativeness. This plot illustrates the level, slope and curvature of the price informativeness curve, as defined in Definition 2.

5 The shape of price informativeness over two decades

For each shock we calculate the three numbers of Definition 2, meaning that we get three time series with as many datapoints as there are shocks in the sample (300). We first study the magnitude and variation of these time series, then how they differ across firm characteristics, and finally we examine how they compare to Ebit-based measures.

5.1 Time-series patterns

We present the output of the recovery procedure of Section 4 in Figure 6. Each panel plots level, slope and curvature of price informativeness (as defined in Definition 2), respectively, over the sample period (1997 through 2022) at the monthly frequency. The purpose of this section is to understand whether their time variation and their magnitude make sense. We start by examining their behavior during the three crises that occurred over our sample: the Dotcom bubble, the global financial crisis of 2008, and the Covid-19 pandemic (all indicated as grey areas in Figure 6). We also examine their fitted trend (dashed red lines) over the sample period. We then discuss magnitudes from the perspective of classical views on market efficiency (Shiller, 1981; Black, 1986). Finally, we verify that these trends cannot be explained by changes in characteristics of shocked firms in our sample (e.g, analyst coverage or price inelasticities) or changes in market conditions (e.g., VIX).

5.1.1 Trends and variation across major events

Unsurprisingly, learning from prices is substantially more difficult during crises. First, level and slope fall sharply on these occasions, with a particularly impressive drop upon the global financial crisis. Therefore, not only does fundamental uncertainty rise during crises, implying there is a larger amount of information to be learnt, but also less information can be learnt from prices over the recovery window. Second, curvature spikes (the information flow becomes strongly convex),

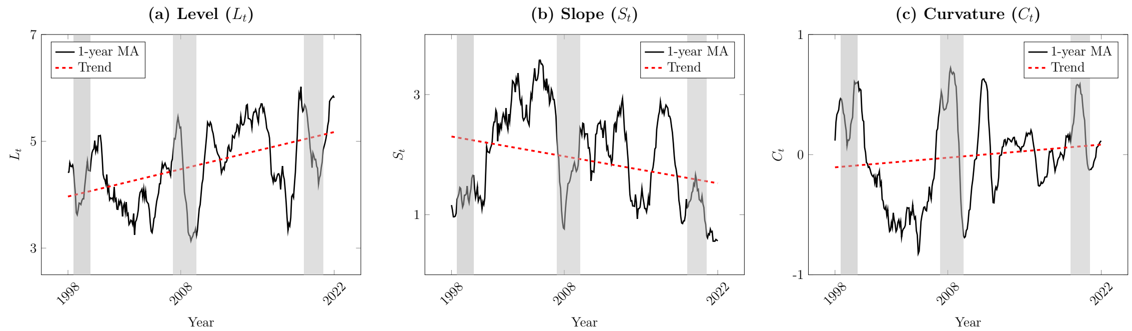

Figure 6: Level, slope and curvature of price informativeness over time. This graph shows the 1-year moving average of the estimated summary statistics of price informativeness as defined in Definition 2, along with a fitted trend line. The shaded areas represent economic crises (Dotcom bubble, global financial crisis and Covid-19 pandemic).

meaning that fundamental information is slower to make its way through prices. Interestingly, towards the end of crises the flow of price informativeness reverses equally fast, with a sudden rebound in level (and slope to some extent) along with a sharp drop in curvature (precision rises, information becomes more abundant and its flow accelerates markedly); this reversal in the shape of price informativeness appears to systematically signal the end of a crisis in our sample.

Two trends are unmistakable. Over the sample period, but mostly over the last decade, the level of price informativeness rose and its slope declined. More specifically, whereas the decline in prior fundamental uncertainty (the rising level) has been quite steady over the sample period (apart from crises episodes) slope rose during the early part of the sample but started declining distinctly after the financial crisis. We will further show that the upward trend in curvature is significant, with curvature swings becoming weaker over time. The picture is quite clear: over the last decade prior uncertainty—the total amount to be learnt about fundamentals—has decreased, and prices have not only revealed less information but have also incorporated this information more slowly.

Level (or prior fundamental precision, τ_F) is particularly important in the analysis because it defines learning potential for a given shock (point in time). Unfortunately, it is the only parameter in $\boldsymbol{\Theta}$ that is unbounded. However, investors' average initial precision, $\tau_F + \phi_0$, is a closely related number for which Theorem 2 provides bounds in (22), which we plot in Figure 7. These bounds (in dashed blue) are quite tight, particularly so in crises, with investors' average precision hitting the lower bound when they occur. These bounds are thus informative, since they provide a reference for what is a high or low level of precision. Given the role τ_F plays in the analysis, it would be helpful to obtain a similar reference for it. We now introduce an argument that will precisely do this, allowing us to appreciate magnitudes (what is a low or high level of price informativeness).

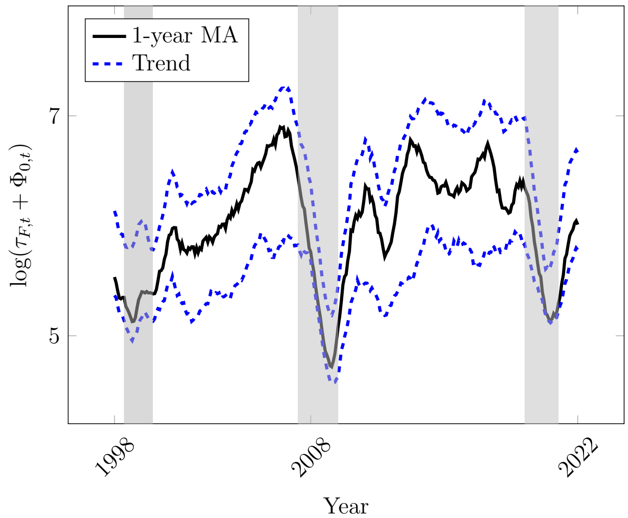

Figure 7: Investors' average initial precision and its bounds. This figure presents the 1-year moving average of investors' average initial precision, the sum of initial fundamental precision and the average initial precision of private information in logs, $\log(\tau_{F,t} + \Phi_{0,t})$. Average initial precision is bounded above and below as per the monotonicity condition (22) in Theorem 2, and these bounds are indicated by dotted blue lines. Shaded gray areas denote economic crises (Dotcom bubble, global financial crisis and Covid-19 pandemic).

5.1.2 Fischer Black's rule of thumb

In his presidential address Fischer Black conjectures "almost all markets are efficient", meaning "price is within a factor 2 of value" at least 90% of the time (Black, 1986). This claim can be formulated mathematically in the model, and we now show it implies an upper bound on fundamental precision, τ_F, along the lines of the classical inequalities formulated in Shiller (1981). We define CAR using the CAR signal in Eq. (15), which takes the intuitive form "value plus noise" consistent with what we think Fischer Black's intuition is:

$$\overline{\mathrm{CAR}}_{0,T} = \frac{\tau_T \cdot P_T - \tau_0 \cdot P_0}{\tau_T - \tau_0} = \widetilde{F} - \frac{a^{1/2}}{a \cdot \tau_T - a \cdot \tau_0} \cdot \int_0^T (1 + a \cdot \phi_t') \cdot \mathrm{d}B_{m,t}. \tag{44}$$

We then compute what Kyle and Obizhaeva (2019) refer to as "pricing accuracy" over the recovery period (in units of noise $1/a$, and using Ito isometry):

$$a \cdot \mathrm{Var}\left[\overline{\mathrm{CAR}}_{0,T} \middle| \widetilde{F}\right]^{-1} = \frac{(a \cdot \tau_T - a \cdot \tau_0)^2}{\int_0^T (1 + a \cdot \phi_t')^2 \mathrm{d}t} \tag{45}$$

$$= \frac{(a \cdot \tau_T - a \cdot \tau_0)^2}{\int_0^T \frac{\exp\left(-2\int_0^t \sigma_u / a^{1/2} du\right)\sigma_t^2 / a}{\left(1/c - \int_0^t \exp\left(-\int_0^v \sigma_u / a^{1/2} du\right)\sigma_v^2 / a dv\right)^2} \mathrm{d}t}, \tag{46}$$

where the second line uses the recovery result of Theorem 2. Since the ratio of two normals is Cauchy distributed the ratio of CAR to value follows:

$$\overline{\text{CAR}}_{0,T}/\widetilde{F} \sim \text{Cauchy}\left(1, \underbrace{\frac{\tau_F^{1/2}}{\text{Var}\left[\overline{\text{CAR}}_{0,T}\middle|\,\widetilde{F}\right]^{-1/2}}}_{\equiv\beta}\right), \tag{47}$$

where β captures the dispersion of the CAR-to-value ratio around 1. Hence, this ratio is "within a factor" x exactly 90% of the time if:

$$\mathbb{P}\left[1/x \leq \overline{\text{CAR}}_{0,T}/\widetilde{F} \leq x \middle| \overline{\text{CAR}}_{0,T}/\widetilde{F} \geq 0\right] = \frac{\frac{1}{\pi}\left(\tan^{-1}\left(\frac{x-1}{\beta}\right) - \tan^{-1}\left(\frac{1/x-1}{\beta}\right)\right)}{1/2 - \tan^{-1}\left(-\frac{1}{\beta}\right)} \tag{48}$$

$$\equiv 90\%. \tag{49}$$

Note that for "within a factor of value" to be meaningful CAR and value must have the same sign, which in this Gaussian framework is not necessarily the case. That is, the ratio of the two must always be positive, hence the conditioning in (48) above. Alternatively, if one thinks of $\overline{\text{CAR}}_{0,T}$ as log-returns and $\widetilde{F}$ as log-fundamentals, the ratio of the two is the log-difference between the two, and thus retains normality (Kyle and Obizhaeva, 2019):

$$\mathbb{P}\left[\frac{1}{\log(x)} \leq \overline{\text{CAR}}_{0,T} - \widetilde{F} \leq \log(x)\right] = \Phi\left(\frac{\log(x)}{\mathbb{V}\left[\overline{\text{CAR}}_{0,T}\middle|\,\widetilde{F}\right]^{1/2}}\right) - \Phi\left(\frac{-\log(x)}{\mathbb{V}\left[\overline{\text{CAR}}_{0,T}\middle|\,\widetilde{F}\right]^{1/2}}\right), \tag{50}$$

where $\Phi(\cdot)$ denotes the standard normal CDF.

Intuitively, the first equation (48) constrains how big the dispersion of prices around value can be and, with a factor $x \equiv 2$, is equivalent to $\beta \approx 0.1541$. Hence, for the CAR-to-value ratio to be within a factor 2 *at least* 90% of the time the following inequality must hold:

$$\tau_F \leq 0.024 \cdot \text{Var}\left[\overline{\text{CAR}}_{0,T}\middle|\,\widetilde{F}\right]^{-1}, \tag{B1}$$

which corresponds to Fischer Black's conjecture in our model. The second equation (50) instead constrains the log-distance between CAR and value, and thus yields an inequality on pricing accuracy in absolute terms (also assuming $x \equiv 2$):

$$1.64/\log(2) \leq \text{Var}\left[\overline{\text{CAR}}_{0,T}\middle|\,\widetilde{F}\right]^{-1/2}. \tag{B2}$$

This inequality (B2) represents the natural interpretation of market efficiency in a CARA-normal context, yet it is unrelated to fundamental precision, τ_F. In contrast, the other inequality (B1) is similar to those of Shiller (1981), in the sense that market efficiency requires fundamental precision,

τ_F, and thus level to be within some fraction of pricing accuracy. The difference is that this fraction in the model reflects a "rule of thumb" (namely, the dispersion parameter of a Cauchy so that 90% of its mass is within a factor 2 of its center), whereas this fraction in Shiller (1981) results from proper optimization (namely, minimizing pricing accuracy for a given level of τ_F).

Since "the factor of 2 is arbitrary, of course" (Black, 1986) the conjecture really hinges on the frequency of inefficiencies. Therefore, we conduct the following, additional exercise: we use our recovered estimate of β and solve equation (48) for x instead. That is, we compute the CAR-to-value factor x necessary for markets to be efficient 90% of the time. We plot this factor (the solution to (48)) in the right panel of Figure 8, and the two versions (B1) and (B2) of Fischer Black's upper bound in the left and middle panel, respectively. Shaded red areas denote times at which Fischer Black would deem markets inefficient.

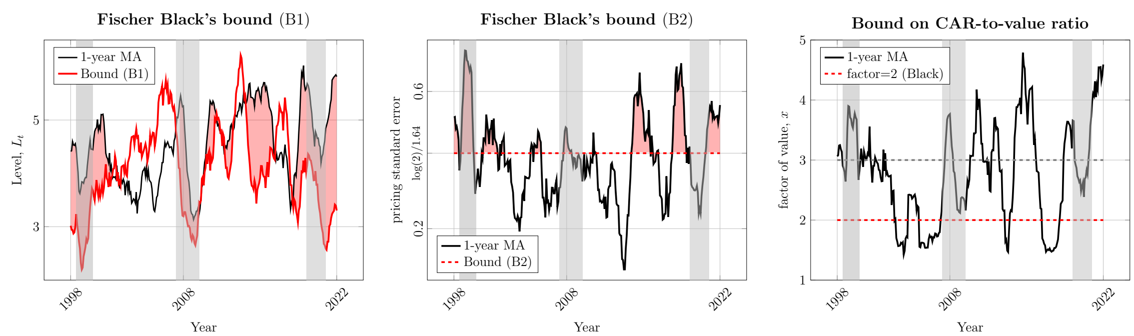

Figure 8: Fischer Black's upper bound and bound on CAR-to-value ratio. The left panel compares level, L, of Definition 2 (in black) to Fischer Black's bound (B1) (in red). The middle panel compares standard pricing error (in black) to Fischer Black's bound (B2) (in red). The right panel compares the minimum CAR-to-value ratio necessary to make Fischer Black's conjecture that the market is efficient 90% of the time true as per (48), and compares it to a factor 2 as proposed by Fischer Black. Shaded red areas highlight periods during which markets become inefficient under each criterion. The gray shaded areas represent economic crises (Dotcom bubble, global financial crisis and Covid-19 pandemic).

We can now better interpret the upward trend in fundamental precision, τ_F, and its magnitude. Focus first on the right panel in Figure 8, which plots the CAR-to-value factor for which the conjecture that markets are efficient 90% of the time is true under the factual dispersion estimate, β. In our sample a factor 2 (lower dashed line) is too optimistic, which explains why periods of inefficiencies (the shaded red areas in the other two panels) occur frequently. A factor 3 (upper dashed line) appears better aligned with the conjecture, which is unsurprising if we consider that firms in our sample presumably experienced a large supply shock and thus that their value should be particularly far from their price. Looking now at the left panel we first confirm the intuition that crises are systematically associated with inefficient markets. This intuition is also verified under the other bound formulation (B2) (middle panel) during the first two crises. Furthermore, under criterion (B1) these inefficiencies were confined to the global financial crisis but precede and extend beyond the Dotcom bubble and, most notably, the pandemic.

What stands out is the long period of substantial inefficiency surrounding the Covid-19 pandemic (under both criteria (B1) and (B2)). This fact suggests that over the last five years of the sample the decline in both fundamental uncertainty (or rise in level) and slope in fact result from inefficient prices. Perhaps intuition would suggest otherwise, namely that prices are more efficient when fundamental precision is higher. Yet, this intuition ignores Shiller (1981)'s insight: what matters is how fundamental precision rises *relative to pricing accuracy*, which is the meaning of the inequality in (B1) (and absent in (B2)). We conclude that the rise in level and the decline in slope over the sample period do not translate in an equivalent improvement in pricing accuracy, especially so in recent years.

A stronger definition of efficiency, different from that of Shiller (1981) and Black (1986), requires that prices instantly reflect all available information (Fama, 1970; LeRoy, 1989). Prices are then efficient if they follow a martingale. In our model the martingale hypothesis fails because price increments in (19) exhibit a trend that even the empircist can predict. However, our notion of curvature is tightly linked to this definition: if we accept the idea that prices do not reflect information instantly but appropriately fast, then this is what curvature precisely captures. The upward trend in the right panel of Figure 6 then confirms our previous conclusion—prices have incorporated information significantly more slowly over the last two decades.

5.1.3 Trends and changes in the characteristics of shocked firms

We are concerned that changes in characteristics of shocked firms may explain the trends we describe. Our sample is arguably specific in that it involves only a fraction of stocks in the market that experience a liquidity shock, whereas evaluating aggregate level and variation typically entails looking at the entire CRSP universe. We now want to verify that trends in the shape of price informativeness are not explained by changes in the composition of the sample of firms we select. For instance, these firms may have become smaller or less liquid. Furthermore, because our empirical framework is based on supply shocks (MFFlow), the patterns we describe could be related to patterns in price inelasticity (Koijen and Yogo, 2019), which certainly vary over time; they are also likely related to information production, e.g., analysts coverage, and to aggregate uncertainty (VIX).

We first examine how our sample of shocked firms differs from the average firm in the NYSE, AMEX and Nasdaq stock universe along traditional characteristics, and thus how representative it is for the whole stock market. In Table 1 we present the mean and median difference between shocked firms and the full sample across *Size*, *Value*, *Liquidity* and analyst *Coverage*. Formally, *Size* and $Value$ are the log of market capitalization and book-to-market ratio, respectively, with book values calculated following Davis, Fama, and French (2000); *Liquidity* is minus Amihud (2002)'s measure of illiquidity; analyst *Coverage* is constructed from I/B/E/S unadjusted summary file by summing the number of forecasts for a given stock across all fiscal periods.[15] In addition, since liquidity is strongly correlated with size we also report it orthogonalized to size, $Liquidity^{\perp}$. Similarly, following Lee and So (2017) we report analyst coverage, $Coverage^{\perp}$, *orthogonalized* to size, turnover and past returns, with the intention of removing the component of coverage attributed to these variables.[16]

	$Size$	$Value$	$Liquidity$	$Liquidity^{\perp}$	$Coverage$	$Coverage^{\perp}$
Δmean	-0.15***	0.10***	1.92***	2.00***	-0.06	0.05*
	(-2.98)	(6.24)	(8.69)	(9.15)	(-1.42)	(1.80)
Δmedian	-0.10*	0.09***	-0.03***	0.66***	-0.12**	0.14***
	(-1.78)	(6.02)	(-4.85)	(5.41)	(-2.44)	(5.31)

Table 1: Summary statistics. This table presents average differences in mean and median of monthly cross-sectional distributions of firm characteristics between the sample of shocked stocks and the full stock universe consisting of US common stocks traded on NASDAQ, AMEX and NYSE excluding shocked stocks. Selected characteristics include $Size$ and $Value$ (log of market cap and book-to-market ratio), *Liquidity* (minus Amihud (2002)'s illiquidity measure) and analyst *Coverage* (log of one plus the sum of forecasts for a given stock in the month in which the shock occurs), with $Coverage^{\perp}$ denoting abnormal $Coverage$ following (Lee and So, 2017) and $Liquidity^{\perp}$ liquidity orthogonalized to $Size$. Book values are calculated following Davis et al. (2000). $Value$, $Size$, and *Liquidity* are calculated at the end of fiscal year t and are valid from July $t+1$ through June $t+2$. The sample spans 300 months, from 1997 through 2021. T-stats are based on standard errors adjusted using Newey and West (1987) with 5 lags, and are shown in parentheses. *, **, and *** denote significance levels at 10%, 5%, and 1%, respectively.

Table 1 confirms the (known) fact that firms mutual funds hold tend to be smaller stocks (Berger, 2023; Wardlaw, 2020). They have higher book-to-market ratios relative to other firms and, although median *Coverage* suggests they receive less attention from analysts, they in fact receive higher coverage once we account for the relatively smaller size of shocked firms ($Coverage^{\perp}$). Similarly, when comparing median with mean *Liquidity* it seems higher liquidity is due to outliers, yet after controlling for size both median and mean $Liquidity^{\perp}$ indicate shocked firms are unambiguously more liquid. Importantly, this result suggests that slow recovery in CARs following a shock (e.g., as in Figure 1 or Figure 2) is not easily explained by slow-moving capital (Duffie, 2010).

We now test whether these deviations in characteristics relative to the entire CRSP universe affect the trends we plotted in Figure 6 and described in the previous sections. For each of the three

[15] In the I/B/E/S unadjusted summary file, we get the total number of forecasts across all fiscal periods by summing "NUMEST" for each ticker and month ("STATPERS"). Following Lee and So (2017) we take the log of it plus one to obtain $Coverage$.

[16] Turnover is calculated as trading volume scaled by shares outstanding, and past performance is measured as the firm's cumulative market-adjusted return over the past 12 months.

statistics (PI_t) of price informativeness in Definition 2 we estimate the following specification:

$$PI_t = \alpha + \mathit{Trend} \times \frac{t}{T} + \sum_i \mathit{controls}_{i,t} + \epsilon_t, \tag{51}$$

where the ratio, t/T, divides calendar time, in our case the $t-$th month (shock) of the sample, by the total number of shocks, $T = 300$, so that it takes values between 0 and 1, with *Trend* thus measuring the average trend in the relevant statistic of price informativeness. In addition, we include a set of control variables, $\mathit{controls}_i$, which contains those used for the summary statistics of Table 1, namely *Size*, *Value*, $\log(\mathit{Liquidity}^{\perp})$, and $\mathit{Coverage}^{\perp}$.[17] To this list we add the CBOE S&P 500 volatility index (VIX) as a measure of aggregate uncertainty, LIQ^{PS} from Pástor and Stambaugh (2003) as a market-wide liquidity measure, and a measure of price inelasticity constructed following van der Beck (2022), all of which represent additional variables that may influence trends.[18] We report estimates of regression (51) in Table 2 for the three statistics of price informativeness.

The main message is, not only do trends persist when including the set of controls (comparing the left and right *Trend* estimate in each column) but they even become statistically stronger. For instance, the trend in curvature in Figure 6 is insignificant but becomes strongly significant when accounting for control variables. This result seems to suggest that the particularities of our sample do not drive our main conclusions regarding trends. Yet, although not a determinant of trends, some control variables are strongly related to price informativeness, among which size and orthogonalized liquidity stand out (we will see this in more details in the next section). In contrast, others appear largely unrelated. For instance, price inelasticities—the strength of the relation between nonfundamental price changes and investor demand—is directly linked to the magnitude of the shock to prices. Although the shock itself does not directly enter our estimation (we only consider the recovery period following it), its magnitude could indirectly affect our estimates. Yet, price inelasticities appear unrelated to our findings. In unreported results we also verify that the magnitude of the shock, the product of price inelasticities with the intensity of fire sales, is equally unrelated.

Given the strong relation of certain firm characteristics with price informativeness in the *time series*, these characteristics could play an important role for price informativeness in *the cross section* of firms, which is what we look at next.

5.2 Cross-sectional patterns

To examine differences in the shape of price informativeness across stocks, we form subgroups of firms along a given characteristic. Specifically, we conduct univariate portfolio sorts on each of

[17] We use the log of *Liquidity* to mitigate the effect of outliers in the regression.

[18] We measure monthly price inelasticity of shocked stocks by estimating the slope coefficients of cross-sectional regressions of abnormal returns on MFflow. We compute abnormal returns relative to the Carhart (1997) factors and industry portfolios (health, manufacturing, hitech and consumer goods), with exposures estimated based on trailing 60 monthly observations.

	Level (L_t)		Slope (S_t)		Curvature (C_t)	
$Const$	3.92***	3.91***	2.33***	3.36***	-0.11	-0.33
	(19.89)	(10.73)	(8.46)	(6.39)	(-0.86)	(-1.42)
$Trend$	1.26***	1.88***	-0.81*	-1.9***	0.2	0.67***
	(3.53)	(4.51)	(-1.84)	(-3.24)	(1.04)	(2.64)
$Size$		0.59***		-0.99***		0.18
		(3.19)		(-4.29)		(1.72)
$Value$		0.47		-0.0		0.21
		(0.72)		(-0.0)		(0.91)
$log(Liquidity^{\perp})$		-0.30		1.42***		0.89***
		(-0.64)		(2.41)		(2.85)
$Coverage^{\perp}$		1.1*		-0.81		0.98***
		(1.86)		(-1.05)		(3.17)
$Inelasticity$		-0.2		0.26		0.15
		(-1.33)		(0.81)		(0.96)
LIQ^{PS}		0.28		1.82*		0.84
		(0.27)		(1.62)		(1.13)
VIX		-0.02		-0.02		-0.01*
		(-1.56)		(-1.4)		(-1.89)

Table 2: Firm characteristics and aggregate price informativeness. This table shows parameter estimates of regression (51). The set of controls includes *Size*, *Value*, (log of) $Liquidity^{\perp}$ as described in Table 1, along with the liquidity measure of Pástor and Stambaugh (2003), a measure of price elasticity constructed following van der Beck (2022) and VIX as a measure of uncertainty. T-statistics are shown in parentheses and are made robust against autocorrelation and heteroskedasticity following (Newey and West, 1987) by using 5 lags. *, **, *** denote the 10%, 5%, and 1% significance levels, respectively.

the characteristics listed in Table 1 ($Size$, $Value$, $Liquidity^{\perp}$ and $Coverage^{\perp}$). Using the median as a breakpoint we form two subportfolios for each characteristic, e.g., small and large stocks. We then follow the same procedure as for the whole sample: for each of these subportfolios we compute *separately* CAR and QV, and apply the recovery procedure of Section 4. Table 3 reports the differences in the three statistics of Definition 2 between large and small, growth and value, high- and low-liquidity and coverage firms, along with their corresponding t-statistic.

	Size	Value	Liquidity$^{\perp}$	Coverage$^{\perp}$
L_t	1.22***	-0.27**	-0.97***	-0.20
	(9.22)	(-2.46)	(-8.13)	(-1.49)
S_t	-0.84***	0.40***	0.62***	0.04
	(-5.28)	(3.04)	(4.30)	(0.33)
C_t	0.13***	-0.06	-0.10**	-0.01
	(2.64)	(-1.54)	(-2.19)	(-0.41)

Table 3: Cross-sectional differences in price informativeness. This table presents the average differences between the three statistics of price informativeness (see Definition 2) across the following characteristics: Size, Value, Liquidity$^{\perp}$ and Coverage$^{\perp}$. The average is calculated from the differences in these statistics between two portfolios that are formed using the cross-sectional median of each characteristic as the breakpoint. These portfolios are created separately for each recovery period. T-statistics are shown in parentheses and are made robust against autocorrelation and heteroskedasticity following (Newey and West, 1987) by using 5 lags. *, **, *** denote the 10%, 5%, and 1% significance levels, respectively.

To a large extent, size and (orthogonalized) liquidity are the two characteristics that matter most for the shape of price informativeness *across firms*, adding to our conclusions regarding the time-series patterns of Table 2. Because larger firms exhibit lower fundamental uncertainty (higher level) they offer less potential for learning and less information to be learnt from their price (lower slope). More surprisingly the flow of information is slower (higher curvature) for larger firms. *After controlling for size* liquidity exhibits exact opposite patterns relative to size. This result is intuitive regarding slope and curvature: higher liquidity is associated with a larger and faster amount of information. At first the result that more liquid firms are associated with greater fundamental uncertainty (lower level) is puzzling. Liquidity in the model is measured by (minus the inverse of) the price coefficient on the supply (e.g., He and Wang (1995), p. 938), which in our case is $-1/\lambda_2 \equiv a^{1/2} \cdot \tau$, and since average precision τ falls with fundamental uncertainty so does liquidity. Yet there is an offsetting effect, which dominates: more liquid firms are also subject to tremendously less noise trading, $1/a$, (the gap in a between high- and low-liquidity firms is 0.47 with a t-statistic of 11.13), which translates directly (and also indirectly through τ) into higher liquidity indeed. Value also influences slope, with high book-to-market firms offering more to learn from prices (higher slope). Yet, our interpretation is that size and liquidity are the two key determinants of the shape of price informativeness across stocks.

What we find most intriguing is that size and liquidity are in fact of vanishing importance for the shape of price informativeness over the sample period. This conclusion is clearly apparent in Figure 9, which plots the difference in level (black), slope (red) and curvature (blue) over the sample period between large and small firms (left panel) and liquid and illiquid firms (right panel), with shaded areas denoting 95% confidence intervals. The shape of price informativeness has become insensitive to size, which is also true for liquidity apart from the difference in level that somewhat persists. Interestingly, this result may speak to the "data feedback loop" (e.g., Begenau et al. (2018) or Veldkamp (2023)), which postulates that large firms benefit more from data, generate more data and thus grow even larger. The trends in Figure 9, which concentrate in the last decade, suggest this mechanism may have become less prevalent among shocked firms.

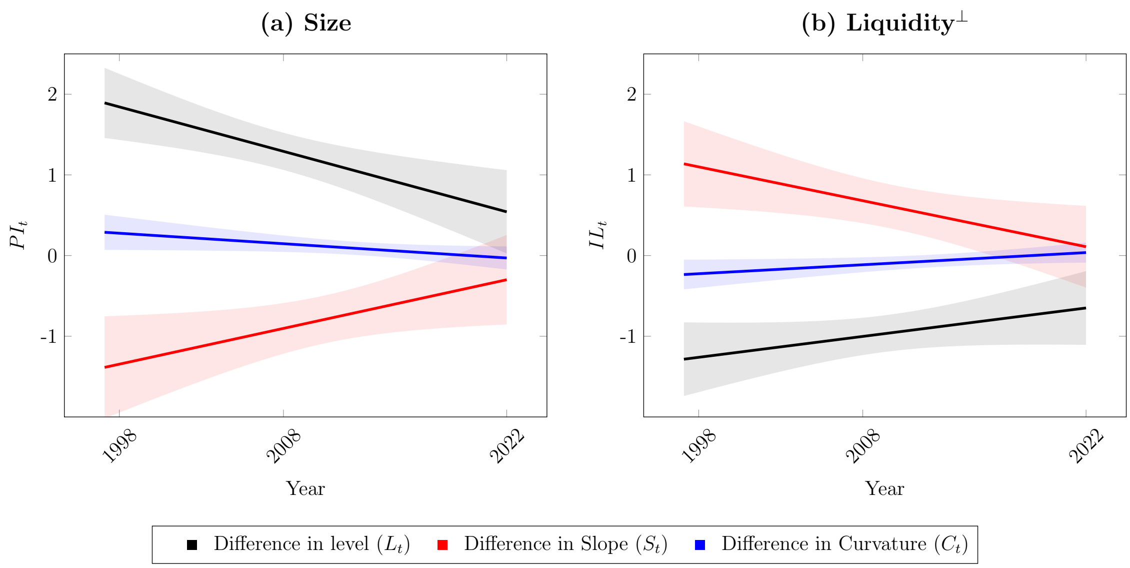

Figure 9: Trends in cross-sectional differences. This figure plots the trends in cross-sectional differences between large and small firms ($Size$), as well as (orthogonalized) liquid and illiquid firms ($Liquidity^{\perp}$) across the three statistics: level (in black), slope (in red) and curvature (in blue). These trends are computed from regressing the differences in these three statistics between the two groups on a monthly normalized calendar variable. The 95% confidence intervals are shown as shaded areas around the estimated trendline. Standard errors are adjusted using the Newey-West method with 5 lags (Newey and West, 1987).

5.3 Comparison with Ebit-based regressions

A first difference between the procedure we propose and Ebit-based regressions is the choice of proxy. We proxy for noise using supply shocks, as opposed to value using earnings, with the advantage of its horizon independence but at the cost of focusing on a subset of shocked firms. Since Ebit-based regressions speak to the substantially broader CRSP universe, our first objective is to understand how these regressions behave *within our sample of shocked firms*. A second difference is that we use intraday data to recover curves of price informativeness, whereas Ebit-based regressions use quarterly data to produce a statistic of price informativeness that is related to their R^2. Therefore, our second objective is to understand how this statistic relates to our curves of price informativeness.

The regression approach regresses future Ebit on market prices. In the literature this regression either exploits cross-sectional variation (Bai et al., 2016; Farboodi et al., 2022) or time-series variation (Davila and Parlatore, 2023). Since our own procedure relies on time-series variation we follow the latter approach, henceforth DP. Our empirical framework identifies a fire sales event, following which a gap between value and prices supposedly forms, and that the market subsequently closes. DP identify this gap in terms of how quarterly log-returns, Δp_t^i, on stock i at date t respond to a change in Ebit over the next x quarters (usually 4, or 1 year):

$$\Delta p_t^i = \beta_0 + \beta_1 \Delta y_{t+x}^i + e_t, \tag{52}$$

where Δy_{t+x}^i is computed as the log of 1 plus the ratio of the absolute change in Ebit and current book value on stock i. Thus the extent to which the market closes the gap is the associated R^2, the fraction of variation in log returns that Δy_{t+x}^i explains, and which we refer to as R^{DP}.

There are two differences in the way we implement this regression relative to DP. We do not account for public information other than returns (e.g., past earnings or profitability), as these controls are absent in our methodology. Yet, following our own approach and consistent with that in Farboodi et al. (2022) we "strip out" common factors in returns, that is we compute Δp_t^i following the abnormal returns construction of Section 2.3 in excess of the Carhart (1997) factors and five-industry portfolios (at the quarterly frequency). Similarly, we remove the aggregate component in in Ebit growth, meaning that we focus on firm-specific cashflows, following Farboodi et al. (2022) who define market Ebit growth as the average across S&P 500 firms, then regress individual Ebit growth on it (and a constant) and track residuals.

A remaining difficulty is to synchronize the regression in (52) with the yearly recovery window of our own procedure. The problem is tied to the quarterly sampling frequency of Ebit, which necessitates a longer window than just a year (4 datapoints at the quarterly frequency) to estimate this regression. Therefore, to these 4 datapoints, we append a window that looks back into the 6.25 years preceding each shock, thus accounting together with the recovery window that follows the shock for a total of 30 quarters (observations) for each shock, a minimum to achieve reliable estimates. The use of a lookback window is not ideal yet necessary due to the frequency of Ebit data. Finally, since each regression is run quarterly we pool the monthly shocked stocks in each quarter.

Each quarter t we obtain an estimate, R_t^{DP}, which we can now map into our curves of price informativeness. Formally, in the model the R^2 of regressing fundamentals, $\widetilde{F}$, on the history of CARs, $(P_t)_{t\leq 1/4}$, over the next *quarter* following the shock is:

$$\mathrm{R}^{\text{this paper}} \equiv \frac{\tau_{1/4}^c - \tau_F}{\tau_{1/4}^c} \tag{53}$$

$$\approx \frac{\exp(S)}{\exp(L) + \exp(S)}, \tag{54}$$

where the first equality corresponds to DP's equation (5), and where we use a quarterly window, as opposed to the whole recovery window, to match the quarterly variation the regression in (52) exploits. The second equality shows that DP's measure summarizes a curve of price informativeness with the ratio of its slope to its terminal point. This measure is thus related to level and slope but separate from curvature, how quickly information becomes available. Figure 10 plots R^{DP} averaged across shocked firms (left panel) and $\mathrm{R}^{\text{this paper}}$ (right panel) over the sample period.

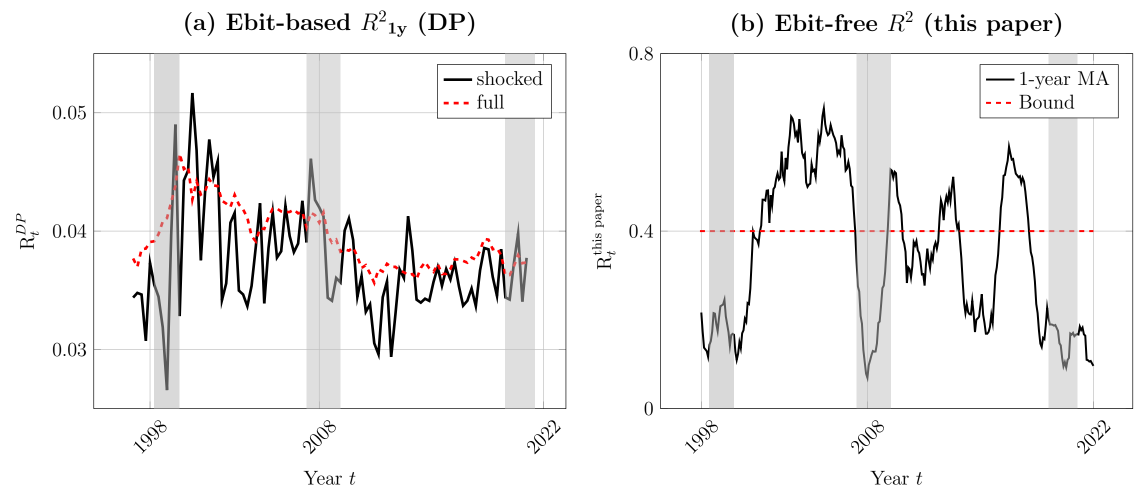

Figure 10: Comparison with Ebit-based measure (DP). Panel (a) plots the R^2 of the Ebit-based regression in (52) over 30 quarters (observations), ranging from the 6.25 years preceding each shock through the 4 quarters following it, which we average across shocked stocks. The independent variable in this regression is residualized one-year-ahead Ebit growth (Δy^i_{t+4}), calculated as the log of one plus the ratio of absolute Ebit change over the next year and current book value. The dependent variable, Δp^i_t, represents quarterly abnormal log-return, in excess of the 4 Carhart (1997) factors and the 5 industry portfolios. Residual Ebit growth is calculated following the construction in Farboodi et al. (2022) (i.e., regressing individual Ebit growth on market Ebit growth and a constant and recording residuals). Panel (b) plots the 1-year moving average of the R^2 in (53) implied by our method over the quarter following the shock. The shaded areas represent economic crises (Dotcom bubble, global financial crisis and Covid-19 pandemic).

Focus first on the left panel, which plots DP's R^2 both within the sample of shocked firms (black solid line) and within the entire CRSP universe (red dashed line). The first conclusion is that the sample of shocked firms, albeit noisier as it involves substantially fewer firms and the identity of which changes from one quarter to the next, is quite representative of the CRSP sample in terms of this R^2. Now looking at the right panel (our own R^2), a clear difference in magnitude and variation is apparent. We examine these two aspects in turn.

Of course, large differences in the magnitude of R^2 are common in different asset-pricing contexts (e.g., Roll (1988), Morck et al. (2000) or Cochrane (2011)). Therefore, it is helpful to provide a model-implied reference for the kind of magnitude we could expect in this specific context. Following similar computations as those of Section 5.1.2 empiricist's inferences relative to fundamentals follow:

$$\widehat{F}^c_{1/4}/\widetilde{F} \sim \text{Cauchy}\left(\mathrm{R}^{\text{this paper}}, \sqrt{\mathrm{R}^{\text{this paper}}(1-\mathrm{R}^{\text{this paper}})}\right). \tag{55}$$

The ratio of the two is precisely centered on the R^2 in (53) and this R^2 determines fluctuations around it, too. We would then expect this ratio to be reasonably close to R^2, and applying Fischer Black's conjecture to it, this would happen for $R^2 \geq 0.4$, the red dashed line in the right panel, and a ten-fold magnitude of DP's R^2. While Ebit did vary over our sample period, it did not vary enough or far enough relative to returns, which could in part explain this difference. The model also abstracts away from elements that could reduce this R^2, e.g., systematic behavioral biases away from fundamentals or residual uncertainty (unlearnable parts in payoffs).

The difference in variation is intuitive, considering that Ebit-based regressions are intended to capture low-frequency information. Unsurprisingly, DP's R^2 is remarkably stable. For instance, our measure indicates that crises (shaded areas) are times of substantial drop in price informativeness, yet such drops are virtually absent in the left panel. This smooth and stable behavior is likely due to the lookback window we impose in our synchronization approach, as it implies a long-weighted moving average of past returns and moving averages tend to smooth the series averaged. The point is, the use of Ebit data, the sampling frequency it imposes and the variation it exhibits relative to returns imply differences in the two approaches and thus in the variation they produce.

Lastly, we examine how Ebit-based measures compare to our results across characteristics. We follow the procedure of Section 5.2, forming two subgroups for each characteristic, and taking the median as breakpoint. We focus on differences in fundamental uncertainty and R^2 in the upper and lower panel of Table 4, respectively. In the case of DP we proxy fundamental uncertainty with the variance of Ebit growth, and report it along with R^2 for different horizons at which we compute Ebit growth, indexed by 1 up to 3 years accordingly. The last line in each panel corresponds to our own measure, $1/\tau_F$ and $\mathrm{R}^{\text{this paper}}$ in the upper and lower panel, respectively.

Although time-series behaviors of the two measures differ largely, they are mostly consistent across characteristics. Focusing on the upper panel first, we see that size and liquidity, the two main cross-sectional determinants of price informativeness in our sample, move the variance of Ebit growth and our measure of fundamental uncertainty in the same direction. They are similarly insensitive to analyst coverage, yet relate to value in opposite directions. Looking now at R^2 in the lower panel both measures of price informativeness appear to move consistently at all horizons, although differences in statistical significance exist. Note that these cross-sectional directions are different from those reported in DP. Farboodi et al. (2022) point out that trends are reverted whether or not controls are included, which we do not. Another reason could be that our sample differs in terms of stock selection (shocked firms) and time period (last two decades). Finally, in unreported results we examine trends in differences across characteristics, and find that differences across large and small firms (and to some extent liquidity) have vanished over the last decade.

Fundamental uncertainty, $\mathbf{Var}[\widetilde{F}]$				
	Size	Value	Liquidity$^{\perp}$	Coverage$^{\perp}$
$Var[\Delta Ebit]_{1y}$	-0.10***	-0.08***	0.12***	0.00
	(-5.74)	(-4.57)	(3.76)	(0.06)
$Var[\Delta Ebit]_{2y}$	-0.14***	-0.14***	0.18***	0.03
	(-4.59)	(-5.35)	(3.48)	(1.13)
$Var[\Delta Ebit]_{3y}$	-0.16***	-0.20***	0.21***	0.06
	(-3.55)	(-5.33)	(2.82)	(1.60)
$1/\tau_F$	-0.14**	0.07**	0.12***	0.03
	(-2.38)	(2.05)	(4.07)	(1.47)
Price informativeness $\mathbf{Var}[\widetilde{F}\|\Delta P]^{-1}$				
	Size	Value	Liquidity$^{\perp}$	Coverage$^{\perp}$
$\mathrm{R}^{\mathrm{DP}}_{1y}$	-0.19	0.33*	0.09	0.64***
	(-1.10)	(1.82)	(0.98)	(3.79)
$\mathrm{R}^{\mathrm{DP}}_{2y}$	-0.27**	0.16	0.21***	0.03
	(-2.53)	(1.16)	(2.79)	(0.33)
$\mathrm{R}^{\mathrm{DP}}_{3y}$	-0.06	0.06	0.04	0.07
	(-0.90)	(0.89)	(0.58)	(0.62)
$\mathrm{R}^{\text{this paper}}$	-0.14***	0.06**	0.11***	0.03
	(-3.95)	(2.18)	(3.34)	(1.11)

Table 4: Comparison of cross-sectional differences with Ebit-based measures. This table shows the average cross-sectional differences in fundamental uncertainty (upper panel) and price informativeness (R^2, lower panel) across $Size$, $Value$, $Liquidity^{\perp}$ and $Coverage^{\perp}$ in the sample of shocked stocks. In the upper panel *Var[*Δ*Ebit]* is the average difference of variance in Ebit growth (dependent variable of regression (52)) between characteristics-based portfolios. In the lower panel R^{DP} corresponds to the difference in the R^2 of regression (52). Indices $1y$, $2y$ and $3y$ correspond to the horizon that is used to calculate Ebit growth. These measures are computed in the sample of shocked stocks over 30 quarters, ranging from 6.25 years prior to each shock through the 4 quarters following it. For ease of comparison with Ebit-based measures, we scale them by a factor 100. The variables $1/\tau_F$ and $\mathrm{R}^{\text{this paper}}$ correspond to our own measure of fundamental uncertainty and R^2 in (53). T-statistics are shown in parentheses and are made robust to autocorrelation and heteroskedasticity following (Newey and West, 1987) by using 5 lags. *, **, *** denote the 10%, 5%, and 1% significance levels, respectively.

References

Acharya, V., H. Almeida, F. Ippolito, and A. Perez (2014). Credit lines as monitored liquidity insurance: Theory and evidence. *Journal of Financial Economics 112*(3), 287–319.

Ali, A., K. D. Wei, and Y. Zhou (2011). Insider trading and option grant timing in response to fire sales (and purchases) of stocks by mutual funds. *Journal of Accounting Research 49*(3), 595–632.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5*(1), 31–56.

Aït-Sahalia, Y., I. Kalnina, and D. Xiu (2020). High-frequency factor models and regressions. *Journal of Econometrics 216*(1), 86–105. Annals Issue in honor of George Tiao: Statistical Learning for Dependent Data.

Bai, J., T. Philippon, and A. Savov (2016). Have financial markets become more informative? *Journal of Financial Economics 122*(3), 625–654.

Barndorff-Nielsen, O. E. and N. Shephard (2004, 01). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics 2*(1), 1–37.

Barndorff-Nielsen, O. E., P. Reinhard Hansen, A. Lunde, and N. Shephard (2009, 11). Realized kernels in practice: trades and quotes. *The Econometrics Journal 12*(3), C1–C32.

Begenau, J., M. Farboodi, and L. Veldkamp (2018). Big data in finance and the growth of large firms. *Journal of Monetary Economics 97*, 71–87.

Berger, E. A. (2023). Selection bias in mutual fund fire sales. *Journal of Financial and Quantitative Analysis 58*(3), 1039–1077.

Black, F. (1986). Noise. *The Journal of Finance 41*(3), 529–543.

Brogaard, J., T. H. Nguyen, T. J. Putnins, and E. Wu (2022, 01). What Moves Stock Prices? The Roles of News, Noise, and Information. *The Review of Financial Studies 35*(9), 4341–4386.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance 52*(1), 57–82.

Cespa, G. and X. Vives (2012). Dynamic trading and asset prices: Keynes vs. hayek. *Review of Economic Studies 79*(2), 539–580.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance 66*(4), 1047–1108.

Coval, J. and E. Stafford (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics 86*(2), 479–512.

Cujean, J. (2020). Idea sharing and the performance of mutual funds. *Journal of Financial Economics 135*(1), 88–119.

Davila, E. and C. Parlatore (2023). Identifying price informativeness. *Working Paper*.

Davis, J. L., E. F. Fama, and K. R. French (2000). Characteristics, covariances, and average returns: 1929 to 1997. *The Journal of Finance 55*(1), 389–406.

Derrien, F., A. Kecskes, and D. Thesmar (2013, December). Investor Horizons and Corporate Policies. (hal-01069196).

Dessaint, O., T. Foucault, L. Frésard, and A. Matray (2018, 11). Noisy Stock Prices and Corporate Investment. *The Review of Financial Studies 32*(7), 2625–2672.

Diamond, D. W. and R. E. Verrecchia (1981, September). Information aggregation in a noisy rational expectations economy. *Journal of Financial Economics 9*(3), 221–235.

Duffie, D. (2010). Presidential address: Asset price dynamics with slow-moving capital. *The Journal of Finance 65*(4), 1237–1267.

Duffie, D. and Y. Sun (2007). Existence of independent random matching. *The Annals of Applied Probability 17*(1), 386 – 419.

Easley, D., N. M. Kiefer, M. O'Hara, and J. B. Paperman (1996). Liquidity, information, and infrequently traded stocks. *The Journal of Finance 51*(4), 1405–1436.

Edmans, A., I. Goldstein, and W. Jiang (2012). The real effects of financial markets: The impact of prices on takeovers. *The Journal of Finance 67*(3), 933–971.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance 25*(2), 383–417.

Farboodi, M., A. Matray, L. Veldkamp, and V. Venkateswaran (2022, 10). Where Has All the Data Gone? *The Review of Financial Studies 35*(7), 3101–3138.

Farboodi, M. and L. Veldkamp (2017). Long run growth of financial technology. Technical report, National Bureau of Economic Research.

Glosten, L. R. and P. R. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics 14*(1), 71–100.

Grossman, S. J. (1995). Dynamic asset allocation and the informational efficiency of markets. *The Journal of Finance 50*(3), 773–787.

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

Gruber, M. J. (1996). Another puzzle: The growth in actively managed mutual funds. *The Journal of Finance 51*(3), 783–810.

Hau, H. and S. Lai (2013). Real effects of stock underpricing. *Journal of Financial Economics 108*(2), 392–408.

Hayashi, T. and N. Yoshida (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli 11*(2), 359–379.

Hayek, F. A. (1945). The use of knowledge in society. *The American Economic Review 35*(4), 519–530.

He, H. and J. Wang (1995). Differential information and dynamic behavior of stock trading volume. *Review of Financial Studies 8*(4), 919–72.

Hellwig, M. F. (1980, June). On the aggregation of information in competitive markets. *Journal of Economic Theory 22*(3), 477–498.

Honkanen, P. and D. Schmidt (2021, 12). Learning from Noise? Price and Liquidity Spillovers around Mutual Fund Fire Sales. *The Review of Asset Pricing Studies 12*(2), 593–637.

Huang, S., M. C. Ringgenberg, and Z. Zhang (2023). The information in asset fire sales. *Management Science 69*(9), 5066–5086.

Koijen, R. S. J. and M. Yogo (2019). A demand system approach to asset pricing. *Journal of Political Economy 127*(4), 1475–1515.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*(6), 1315–1335.

Kyle, A. S. and A. Obizhaeva (2019). Market microstructure invariance: A dynamic equilibrium model. *Unpublishded Manuscript*.

Lee, C. M. and E. C. So (2017). Uncovering expected returns: Information in analyst coverage proxies. *Journal of Financial Economics 124*(2), 331–348.

LeRoy, S. F. (1989). Efficient capital markets and martingales. *Journal of Economic Literature 27*(4), 1583–1621.

Levene, H. (1960). Robust tests for equality of variances. In *Contributions to probability and statistics*, Volume 2 of *Stanford Studies in Mathematics and Statistics*, pp. 278–292. Stanford Univ. Press, Stanford, CA.

Liptser, R. and A. Shiryaev (2001). *Statistics of Random Processes II.* Springer Verlag, New York.

Mancini, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping brownian motion. *Giornale dell'Istituto Italiano degli Attuari 64*(19-47), 44.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics 36*(2), 270–296.

Milgrom, P. and N. Stokey (1982). Information, trade and common knowledge. *Journal of Economic Theory 26*(1), 17–27.

Morck, R., B. Yeung, and W. Yu (2000). The information content of stock markets: why do emerging markets have synchronous stock price movements? *Journal of Financial Economics 58*(1), 215–260. Special Issue on International Corporate Governance.

Newey, W. K. and K. D. West (1987, May). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–08.

Norli, O., C. Ostergaard, and I. Schindele (2014, 10). Liquidity and Shareholder Activism. *The Review of Financial Studies 28*(2), 486–520.

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political economy 111*(3), 642–685.

Roll, R. (1988). *The Journal of Finance 43*(3), 541–566.

Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *The American Economic Review 71*(3), 421–436.

Sirri, E. R. and P. Tufano (1998). Costly search and mutual fund flows. *The Journal of Finance 53*(5), 1589–1622.

van der Beck, P. (2022, June). On the Estimation of Demand-Based Asset Pricing Models. (22-67).

Veldkamp, L. (2023, 01). Valuing Data as an Asset*. *Review of Finance 27*(5), 1545–1562.

Wang, J. (1994, February). A model of competitive stock trading volume. *Journal of Political Economy 102*(1), 127–68.

Wardlaw, M. (2020). Measuring mutual fund flow pressure as shock to stock returns. *The Journal of Finance 75*(6), 3221–3243.

Weller, B. M. (2018). Does algorithmic trading reduce information acquisition? *The Review of Financial Studies 31*(6), 2184–2226.

Zuo, L. (2016). The informational feedback effect of stock prices on management forecasts. *Journal of Accounting and Economics 61*(2), 391–413.

A Appendix

A.1 Proof of Theorem 1

This appendix summarizes the main steps in Cujean (2020) and proposes an alternative "guess-and-verify" argument regarding the derivation of equilibrium coefficients. The steps consists in "guessing and verifying" that prices have the form:

$$P_t = \lambda_{1,t}\widetilde{F} + (1-\lambda_{1,t})\widehat{F}_t^c + \lambda_{2,t}\widetilde{m}_t, \tag{56}$$

for $t < T$ with $\lambda_{1,t}$ and $\lambda_{2,t}$ deterministic, equilibrium functions.

Step 1. We start by building for each individual agent the updating rule piecewise. Let $\mathbf{x}_t \equiv (\widetilde{F}, \widetilde{m}_t)'$ be the vector of unobservables with dynamics:

$$\mathrm{d}\mathbf{x}_t = \begin{pmatrix} 0 & 0 \\ 0 & -b_m \end{pmatrix} \mathbf{x}_t \mathrm{d}t + \begin{pmatrix} 0 \\ \tau_m^{-1/2} \end{pmatrix} \mathrm{d}B_{m,t} \equiv \mathbf{a}\mathbf{x}_t \mathrm{d}t + \mathbf{b}\mathrm{d}B_{m,t}, \tag{57}$$

and let:

$$\{0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq ... \leq \tau_{N_T^i} \leq \tau_{N_T^i+1} = T\} \tag{58}$$

be Poisson arrival times at which investor i receives new private signals, meaning activation times of the counter N^i. Consider first dates $t \in (\tau_k, \tau_{k+1})$ for $k \in \{0, ..., N_T^i\}$ when new information flows from prices exclusively. From the conjecture in (56) and because all agents observe empiricist's information, $\xi_t \equiv P_t - (1-\lambda_{1,t})\widehat{F}_t^c = \lambda_{1,t}\widetilde{F} + \lambda_{2,t}\widetilde{m}_t$ is a sufficient statistic for prices, and differentiating:

$$\mathrm{d}\xi_t = \begin{pmatrix} \lambda'_{1,t} & \lambda'_{2,t} - b_m\lambda_{2,t} \end{pmatrix} \mathbf{x}_t \mathrm{d}t + \lambda_{2,t}\tau_m^{-1/2}\mathrm{d}B_{m,t} \equiv \mathbf{A}_{1,t}\mathbf{x}_t \mathrm{d}t + \mathbf{B}_{1,t}\mathrm{d}B_{m,t}. \tag{59}$$

Since this information is continuous posterior mean and variance, $\widehat{\mathbf{x}}_t^i \equiv \mathbb{E}[\mathbf{x}_t|\mathscr{F}_t^i]$ and $\mathbf{O}_t^i \equiv \mathbb{V}[\mathbf{x}_t|\mathscr{F}_t^i]$, are updated according to a continuous-time Kalman filter:

$$\mathrm{d}\widehat{\mathbf{x}}_t^i = \mathbf{a}\widehat{\mathbf{x}}_t^i \mathrm{d}t + (\mathbf{O}_t^i\mathbf{A}'_{1,t} + \mathbf{b}\mathbf{B}'_{1,t})(\mathbf{B}_{1,t}\mathbf{B}'_{1,t})^{-\frac{1}{2}}\mathrm{d}\widehat{B}_t^i \tag{60}$$

$$\dot{\mathbf{O}}_t^i = \mathbf{a}\mathbf{O}_t^i + \mathbf{O}_t^i\mathbf{a}' + \mathbf{b}\mathbf{b}' - (\mathbf{O}_t^i\mathbf{A}'_{1,t} + \mathbf{b}\mathbf{B}'_{1,t})(\mathbf{B}_{1,t}\mathbf{B}'_{1,t})^{-1}(\mathbf{A}_{1,t}\mathbf{O}_t^i + \mathbf{B}'_{1,t}\mathbf{b}), \tag{61}$$

where the filter innovation is:

$$\mathrm{d}\widehat{B}_t^i = (\mathbf{B}_{1,t}\mathbf{B}'_{1,t})^{-\frac{1}{2}}(\mathrm{d}\xi_t - \mathbf{A}_{1,t}\widehat{\mathbf{x}}_t^i \mathrm{d}t). \tag{62}$$

Consider now times $t = \tau_k$ at which agent i receives a chunk of Δn_t^i new signals, for which the average in (6) is a sufficient statistic and which we rewrite as:

$$Y_t^i \equiv \begin{pmatrix} 1 & 0 \end{pmatrix}\mathbf{x}_t + (\Delta n_t^i \tau_S)^{-1/2}\epsilon_t^i \equiv \mathbf{A}_{2,t}\mathbf{x}_t + B_{2,t}(\Delta n_t^i)\epsilon_t^i. \tag{63}$$

The updating rule in this case is a discrete-time Kalman filter:

$$\widehat{\mathbf{x}}_t^i = \widehat{\mathbf{x}}_{t-}^i + \mathbf{O}_{t-}^i\mathbf{A}'_{2,t}\left(\mathbf{A}_{2,t}\mathbf{O}_{t-}^i\mathbf{A}'_{2,t} + B_{2,t}(\Delta n^i)^2\right)^{-1}\widehat{Y}_t^i \equiv \boldsymbol{\omega}_t o_t^i \frac{\Delta n_t^i}{\sigma_S^2}\widehat{Y}_t^i \tag{64}$$

$$\mathbf{O}_t^i = \mathbf{O}_{t-}^i - \mathbf{O}_{t-}^i\mathbf{A}'_{2,t}\left(\mathbf{A}_{2,t}\mathbf{O}_{t-}^i\mathbf{A}'_{2,t} + B_{2,t}(\Delta n^i)^2\right)^{-1}\mathbf{A}_{2,t}\mathbf{O}_{t-}^i \tag{65}$$

where $\boldsymbol{\omega}_t \equiv \begin{pmatrix} 1 & -\lambda_{1,t}/\lambda_{2,t} \end{pmatrix}'$ and the filter innovation is $\widehat{Y}_t^i = Y_t^i - \mathbb{E}\left[Y_t^i|\mathscr{F}_{t-}^i, \Delta n_t^i\right]$. We can then use observational equivalence to rewrite the variance-covariance matrix $\mathbf{O}_t^i$ as:

$$\mathbf{O}_t^i = \begin{pmatrix} 1 & -\lambda_{1,t}/\lambda_{2,t} \\ -\lambda_{1,t}/\lambda_{2,t} & (\lambda_{1,t}/\lambda_{2,t})^2 \end{pmatrix}\frac{1}{\tau_t^i} \equiv \boldsymbol{\Omega}_t \frac{1}{\tau_t^i}, \tag{66}$$

with $\tau_t^i \equiv \mathbb{V}[\widetilde{F}|\mathscr{F}_t^i]$. Using this expression to simplify the updating rules in Eq. (60) and Eq. (64) and putting the two pieces together the updating rule at any time satisfies (omitting initial conditions):

$$\mathrm{d}\widehat{\mathbf{x}}_t^i = \mathbf{a}\widehat{\mathbf{x}}_{t-}^i \mathrm{d}t + \begin{pmatrix} k_t/\tau_{t-}^i \\ \tau_m^{-1/2} - \lambda_{1,t}/\lambda_{2,t}k_t/\tau_{t-}^i \end{pmatrix}\mathrm{d}\widehat{B}_t^i + \boldsymbol{\omega}_t\tau_S\Delta n_t^i/\tau_t^i\widehat{Y}_t^i\mathrm{d}N_t^i \tag{67}$$

$$\mathrm{d}\tau_t^i = k_t^2\mathrm{d}t + \tau_S\Delta n_t^i \mathrm{d}N_t^i, \tag{68}$$

where we define the speed at which prices reveal information as:

$$k_t \equiv \tau_m^{1/2}\left(\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\lambda_{1,t}}{\lambda_{2,t}}\right) + b_m\frac{\lambda_{1,t}}{\lambda_{2,t}}\right). \tag{69}$$

Step 2. Defining $\Delta^i \equiv \widehat{F}^i - \widehat{F}^c$ we show that $\boldsymbol{\Psi}^i \equiv \begin{pmatrix} \Delta^i & \widehat{m}^i \end{pmatrix}'$ is Markovian and fully determines the way agent i perceives her investment opportunity set. Because the empiricist only observes information from prices her views satisfy:

$$\mathrm{d}\widehat{\mathbf{x}}_t^c = \mathbf{a}\mathbf{x}_t^c \mathrm{d}t + \begin{pmatrix} k_t/\tau_t^c \\ \tau_m^{-1/2} - \lambda_{1,t}/\lambda_{2,t} k_t/\tau_t^c \end{pmatrix} \mathrm{d}\widehat{B}_t^c \tag{70}$$

$$\mathrm{d}\tau_t^c = k_t^2 \mathrm{d}t, \tag{71}$$

where the filter innovation $\widehat{B}_t^c$ under $\mathscr{F}^c$ (defined in (10)) satisfies:

$$\mathrm{d}\widehat{B}_t^c = (\mathbf{B}_{1,t}\mathbf{B}_{1,t}')^{-\frac{1}{2}}(\mathrm{d}\xi_t - \mathbf{A}_{1,t}\widehat{\mathbf{x}}_t^c \mathrm{d}t). \tag{72}$$

Define the change of probability measure between $\widehat{\mathbb{P}}^c$ and $\widehat{\mathbb{P}}^i$ under $\mathscr{F}^i$ as:

$$Z_t = \left.\frac{\mathrm{d}\widehat{\mathbb{P}}^i}{\mathrm{d}\widehat{\mathbb{P}}^c}\right|_{\mathscr{F}_t^i} = \exp\left(-\frac{1}{2}\int_0^t (k_s\Delta_s^i)^2 \mathrm{d}s + \int_0^t k_s \Delta_s^i \mathrm{d}\widehat{B}_s^c\right), \tag{73}$$

with the two innovations $\widehat{B}^i$ and $\widehat{B}^c$ related as:

$$\widehat{B}_t^i = \widehat{B}_t^c - \int_0^t k_s \Delta_s^i \mathrm{d}s, \tag{74}$$

and the precisions $\tau_t^i \equiv \tau_t(n^i)$ and τ_t^c related as:

$$\tau_t(n^i) = \mathrm{Var}\left[\widetilde{F}\middle| \mathscr{F}_t^i; n_t^i = n^i\right]^{-1} = \tau_t^c + \frac{n^i}{\sigma_S^2}, \tag{75}$$

$$\tau_t^c = \tau_F + \int_0^t k_s^2 \mathrm{d}s. \tag{76}$$

Using (74) to express the dynamics of $\widehat{F}^c$ under $\widehat{\mathbb{P}}^i$ and rearranging gives:

$$\mathrm{d}\boldsymbol{\Psi}_t^i = \begin{pmatrix} -k_t^2/\tau_t^c & 0 \\ 0 & -b_m \end{pmatrix} \boldsymbol{\Psi}_{t-}^i \mathrm{d}t + \begin{pmatrix} (\frac{1}{\tau_t(n_{t-}^i)} - \frac{1}{\tau_t^c})k_t \\ \tau_m^{-1/2} - \frac{\lambda_{1,t}}{\lambda_{2,t}}\frac{k_t}{\tau_t(n_{t-}^i)} \end{pmatrix} \mathrm{d}\widehat{B}_t^i + \boldsymbol{\omega}_t \frac{\tau_S}{\tau_t(n_t^i)} \Delta n_t^i \widehat{Y}_t^i \mathrm{d}N_t^i \tag{77}$$

$$\equiv \mathbf{A}_{\Psi,t}\boldsymbol{\Psi}_{t-}^i \mathrm{d}t + \mathbf{B}_{\Psi,t}(n_{t-}^i)\mathrm{d}\widehat{B}_t^i + \mathbf{C}_{\Psi,t}(n_{t-}^i, \Delta n_t^i)\widehat{Y}_t^i \mathrm{d}N_t^i, \tag{78}$$

which shows that $\boldsymbol{\Psi}^i$ is Markovian. To show that it determines agent i's investment opportunity set differentiate the conjecture in (56) and substitute the relevant SDEs under $\mathscr{F}^i$.

After rearranging we confirm that price paths are continuous over $t \in (0, T)$ and satisfy:

$$\mathrm{d}P_t = \begin{pmatrix} \lambda_{1,t}' + (1-\lambda_{1,t})\frac{k_t^2}{\tau_t^c} & \lambda_{2,t}' - b_m\lambda_{2,t} \end{pmatrix} \boldsymbol{\Psi}_t^i \mathrm{d}t + (\lambda_{2,t}\tau_m^{-1/2} + (1-\lambda_{1,t})/\tau_t^c k_t)\mathrm{d}\widehat{B}_t^i \tag{79}$$

$$\equiv \mathbf{A}_{P,t}\boldsymbol{\Psi}_t^i \mathrm{d}t + B_{P,t}\mathrm{d}\widehat{B}_t^i. \tag{80}$$

At date T, however, prices jump to $\widetilde{F}$. Using the conjecture in (56), under $\mathscr{F}_{T-}^i$ they jump by:

$$\Delta P_T = \Pi - P_{T-} \equiv \begin{pmatrix} 1-\lambda_{1,T-} & -\lambda_{2,T-} \end{pmatrix} \boldsymbol{\Psi}_{T-}^i \equiv \boldsymbol{\lambda}_{T-}' \boldsymbol{\Psi}_{T-}^i, \tag{81}$$

which shows that $\boldsymbol{\Psi}^i$ fully determines agent i's investment opportunity set.

Step 3. Given how agents perceive their investment opportunity set, we now determine their optimal demand, θ^i. Let the value function associated with (2) at time t be:

$$J(W^i, \boldsymbol{\Psi}^i, n^i, t) = \max_{\theta^i} \mathbb{E}\left[-\exp\left(-\gamma W_T^i\right) \middle| \mathscr{F}_t^i; W_t^i = W^i, \boldsymbol{\Psi}_t^i = \boldsymbol{\Psi}^i, n_t^i = n^i \right] \tag{82}$$

$$\text{s.t. } \mathrm{d}W_t^i = \theta_t^i \mathrm{d}P_t. \tag{83}$$

Standard arguments imply J solves the HJB equation (omitting i indices for brevity):

$$0 = \max_{\theta} \left\{ J_W \mathbf{A}_P \boldsymbol{\Psi}\theta + \frac{1}{2} J_{WW} B_P^2 \theta^2 + B_P \mathbf{B}_\Psi(n)' J_{W\Psi}\theta \right\} + \frac{1}{2}\mathrm{tr}(J_{\Psi\Psi}\mathbf{B}_\Psi(n)\mathbf{B}_\Psi(n)') \tag{84}$$
$$+ J_t + J_\Psi' \mathbf{A}_\Psi \boldsymbol{\Psi} + \eta(n)\mathbb{E}^{\mathscr{L}_t(\widehat{Y},\Delta n)}\left[J(W, \boldsymbol{\Psi} + \mathbf{C}_\Psi(n,\Delta n)\widehat{Y}, n+\Delta n, t) - J(W, \boldsymbol{\Psi}, n, t) \right],$$

where $\eta(n)$ denotes the intensity at which new signals accrue (as defined in Section 3.2) and $\mathscr{L}_t$ denotes the joint density of the filter innovation $\widehat{Y}_t^i | \mathscr{F}_{t-}^i \sim \mathscr{N}(0, 1/\tau_{t-}^i + (\tau_S \Delta n_t^i)^{-1})$ and $\Delta n_t^i \sim \mu_t$ (with μ satisfying (8)). Given the price discontinuity at time T, the boundary condition is a static optimization problem of the form:

$$J(W, \boldsymbol{\Psi}, n, T-) = \max_{\theta} -\exp\left(-\gamma \left(W + \theta_{T-}\boldsymbol{\lambda}_{T-}'\boldsymbol{\Psi} - \frac{1}{2}\gamma\theta_{T-}^2 \boldsymbol{\lambda}_{T-}' \boldsymbol{\Omega}_{T-}\boldsymbol{\lambda}_{T-}/\tau_{T-} \right)\right). \tag{85}$$

Taking first-order conditions on the HJB equation and the boundary condition gives:

$$\theta_t^i \equiv \theta_t(\boldsymbol{\Psi}^i, n^i) = -\frac{J_W \mathbf{A}_P \boldsymbol{\Psi}^i + B_P \mathbf{B}_\Psi(n^i)' J_{W\Psi}}{J_{WW} B_P^2} \tag{86}$$

$$\theta_{T-}^i = \frac{1}{\gamma}\tau_{T-}(n)\boldsymbol{\lambda}_{T-}^\top \boldsymbol{\Psi}^i. \tag{87}$$

Substituting back into the HJB equation gives (omitting i indices):

$$
\begin{aligned}
0 = {} & J_t + J_\Psi' \mathbf{A}_\Psi \boldsymbol{\Psi} + \frac{1}{2}\mathrm{tr}(J_{\Psi\Psi}\mathbf{B}_\Psi(n)\mathbf{B}_\Psi(n)') - \frac{1}{2}\frac{(J_W \mathbf{A}_P \boldsymbol{\Psi} + B_P \mathbf{B}_\Psi(n)' J_{W\Psi})^2}{J_{WW} B_P^2} \\
& + \eta(n)\mathbb{E}^{\mathscr{L}_t(\widehat{Y},\Delta n)}\left[J(W, \boldsymbol{\Psi} + \boldsymbol{\sigma}(n,\Delta n)\widehat{Y}, n+\Delta n, t) - J(W,\boldsymbol{\Psi}, n, t)\right]
\end{aligned}
\tag{88}
$$

and similarly for the boundary condition:

$$
J(W, \boldsymbol{\Psi}, n, T-) = -\exp\left(-\gamma W - \frac{1}{2}\tau_{T-}(n)\boldsymbol{\Psi}'\boldsymbol{\Lambda}_{T-}\boldsymbol{\Psi}\right). \tag{89}
$$

We conjecture J is of the affine-quadratic form:

$$
J(W, \boldsymbol{\Psi}, n, t) = -\exp\left(-\gamma W - u_t(n) - \frac{1}{2}\boldsymbol{\Psi}'\mathbf{M}_t(n)\boldsymbol{\Psi}\right), \tag{90}
$$

where u and $\mathbf{M}$ are, respectively, scalar and $2\times 2-$matrix coefficients to be determined. Note there are no linear terms in $\boldsymbol{\Psi}$ because states have unconditional mean 0. Note further $u_t(n)$ does not intervene in optimal demands and thus our focus is exclusively on $\mathbf{M}_t(n)$. Whereas the quadratic conjecture can be verified in a diffusive CARA-Gaussian setup it usually fails in a jump-diffusion setup unless jumps are restricted to the linear part (Coval and Stafford, 2007), since the conjecture in (90) would imply the jump term in the HJB equation is:

$$
\begin{aligned}
& \mathbb{E}^{\mathscr{L}_t(\widehat{Y},\Delta n)}\left[\frac{J(W, \boldsymbol{\Psi} + \mathbf{C}_\Psi(n,\Delta n)\widehat{Y}, n+\Delta n, t)}{J(W,\boldsymbol{\Psi},n,t)} - 1\right] \equiv \\
& \sum_{m\in\mathbb{N}} \pi_t(m;n)\frac{e^{u_t(n) - u_t(n+m) - \frac{1}{2}\boldsymbol{\Psi}'((\mathbf{I}+\mathbf{M}_t(n+m)\boldsymbol{\Sigma}_t(n,m))^{-1}\mathbf{M}_t(n+m) - \mathbf{M}_t(n))\boldsymbol{\Psi}}}{|\mathbf{I} + \boldsymbol{\Sigma}_t(n,m)\mathbf{M}_t(n+m)|^{\frac{1}{2}}},
\end{aligned}
\tag{91}
$$

where $\pi(\cdot)$ denotes the distribution of number of new signals (as defined in Section 3.2) and where we have defined:

$$
\boldsymbol{\Sigma}_t(n,m) \equiv \boldsymbol{\Omega}_t \frac{\tau_S m}{\tau_t(n)\tau_t(n+m)}. \tag{92}
$$

However, Cujean (2020) shows $\mathbf{M}$ satisfies the difference equation (in the $n-$dimension):

$$
\mathbf{M}_t(n) = (\mathbf{I} + \mathbf{M}_t(n+m)\boldsymbol{\Sigma}_t(n,m))^{-1}\mathbf{M}_t(n+m), \tag{93}
$$

so that from (91) the jump in the quadratic part (in $\boldsymbol{\Psi}$) is exactly zero and thus restricted to the linear part (which in this case is also 0). As a result and remarkably, the quadratic form in (90) extends to this jump-diffusion context. In particular, plugging (90) in the HJB equation further gives a matrix Riccati equation for $\mathbf{M}$ (in the $t-$dimension):

$$\dot{\mathbf{M}}_t(n) = -\frac{\mathbf{A}'_{P,t}\mathbf{A}_{P,t}}{B^2_{P,t}} + \mathbf{M}_t(n)\left(\frac{\mathbf{B}_{\Psi,t}(n)\mathbf{A}_{P,t}}{B_{P,t}} - \mathbf{A}_{\Psi,t}\right) + \left(\frac{\mathbf{B}_{\Psi,t}(n)\mathbf{A}_{P,t}}{B_{P,t}} - \mathbf{A}_{\Psi,t}\right)'\mathbf{M}_t(n), \tag{94}$$

subject to the boundary condition (by comparing the conjecture with (89)):

$$\mathbf{M}_{T-}(n) = \tau_{T-}(n)\mathbf{\Lambda}_{T-}. \tag{95}$$

Intuitively, the quadratic form persists because *1.* prices have continuous sample paths they do not covary with discontinuous bulks of signals when they arise, meaning that they do not create a hedging demand and *2.* the law of iterated expectations then implies the jump in the value function should be a martingale. This result is the reason for which the dynamics of information collection in Section 3.2 can be quite general (as long as they do not depend on aggregate states). Plugging (90) in the first-order condition gives:

$$\theta_t(\mathbf{\Psi}^i, n^i) = \frac{\mathbf{A}_P - B_P\mathbf{B}_\Psi(n^i)'\mathbf{M}(n^i)}{\gamma B_P^2}\mathbf{\Psi}^i. \tag{96}$$

Step 4. We now clear the stock market:

$$\int_0^1 \theta_t^i \mathrm{d}i = \widetilde{m}_t. \tag{97}$$

To aggregate individual demands use the updating rule to differentiate $\tau^i\widehat{F}^i$ and $\tau^c\widehat{F}^c$, substract one from the other and integrate:

$$\tau_t(n_t^i)\widehat{F}_t^i = \tau_t^c\widehat{F}_t^c + \sum_{s\le t}\tau_S\Delta n_s^i Y_s^i\Delta N_s^i. \tag{98}$$

Then plug the expression for the average private signal Y^i in (6) and rearrange:

$$\widehat{F}_t^i = \frac{\tau_t^c}{\tau_t(n_t^i)}\widehat{F}_t^c + \frac{n_t^i\tau_S}{\tau_t(n_t^i)}\widetilde{F} + \frac{1}{\tau_t(n_t^i)}\sum_{s\le t}\left(\tau_S\Delta n_s^i\right)^{\frac{1}{2}}\epsilon_s^i\Delta N_s^i \tag{99}$$

Aggregating across agents the law of large numbers (namely, that $\int_0^1 \epsilon_s^i \mathrm{d}i = 0$) implies:

$$\int_0^1 \widehat{F}_t^i \mathrm{d}i = \sum_{n\in\mathbb{N}} \mu_t(n)((1-\alpha_t(n))\widehat{F}_t^c + \alpha_t(n)\widetilde{F}), \tag{100}$$

where $\alpha_t(n) \equiv \frac{\tau_t(n)-\tau_t^c}{\tau_t(n)}$ is the weight Bayes' rule assigns to $\widetilde{F}$ in consensus beliefs. Using this relation, the definition of $\mathbf{\Psi}^i$ and observational equivalence we conclude that:

$$\int_0^1 \mathbf{\Psi}_t^i \mathrm{d}i = \sum_{n\in\mathbb{N}} \mu_t(n) \begin{pmatrix} \alpha_t(n) & 0 \\ \frac{\lambda_{1,t}}{\lambda_{2,t}}(1-\alpha_t(n)) & 1 \end{pmatrix} \mathbf{\Psi}_t, \tag{101}$$

where $\mathbf{\Psi}_t \equiv \begin{pmatrix} \widetilde{F} - \widehat{F}_t^c & \widetilde{m}_t \end{pmatrix}'$. Plugging this aggregation result along with optimal demands in (96) into the market-clearing condition, and separating variables gives a system of two ODEs for the equilibrium price coefficients λ_1 and λ_2 :

$$\sum_{n\in\mathbb{N}} \mu_t(n) \frac{\mathbf{A}_P - B_P \mathbf{B}_\Psi(n)'\mathbf{M}(n)}{\gamma B_P^2} \begin{pmatrix} \alpha_t(n) & 0 \\ \frac{\lambda_{1,t}}{\lambda_{2,t}}(1-\alpha_t(n)) & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \end{pmatrix}. \tag{102}$$

Stock market-clearing at the final date, $\int_0^1 \theta_{T-}^i \mathrm{d}i = \widetilde{m}_T$, gives associated boundary conditions:

$$\begin{pmatrix} \lambda_{1,T-} & \lambda_{2,T-} \end{pmatrix} = \begin{pmatrix} \tau_{T-} - \tau_{T-}^c & -\gamma \end{pmatrix} \frac{1}{\tau_{T-}}, \tag{103}$$

which concludes the equilibrium construction and verifies the initial conjecture.

Step 5. (alternative argument) The difference and differential equations for $\mathbf{M}$ and the two ODEs for λ_1 and λ_2 form a system of coupled equations, the solution of which represents the equilibrium we seek. If we could guess

[TBC]

A.2 Proof of Theorem 2

Define $g_t \equiv 1 + a\phi_t'$ and rewrite the ODE in (20) accordingly:

$$a^{1/2}\tau_t = g_t/\sigma_t. \tag{104}$$

Differentiating this equation once more with respect to time we get:

$$g_t(g_t - 1) = a^{1/2}(g_t/\sigma_t)'. \tag{105}$$

Introducing the following transformation next:

$$g_t = e^{-\int_0^t \sigma_s/a^{1/2}ds}/f_t, \tag{106}$$

gives the following ODE for f:

$$(f_t\sigma_t)' = -e^{-\int_0^t \sigma_s/a^{1/2}ds}\sigma_t^2/a^{1/2}. \tag{107}$$

Integrating and undoing the transformation gives:

$$g_t = \frac{\sigma_t/a^{1/2}e^{-\int_0^t \sigma_s/a^{1/2}\mathrm{d}s}}{\sigma_0/a^{1/2}/g_0 - \int_0^t e^{-\int_0^v \sigma_s/a^{1/2}\mathrm{d}s}\sigma_v^2/a\mathrm{d}v}. \tag{108}$$

Substituting the definition of g_t and integrating we get:

$$\phi_t = \phi_0 + \frac{1}{a}\int_0^t \left(\frac{\sigma_u/a^{1/2}e^{-\int_0^u \sigma_s/a^{1/2}\mathrm{d}s}}{\sigma_0/a^{1/2}/(1+a\phi_0') - \int_0^u e^{-\int_0^v \sigma_s/a^{1/2}\mathrm{d}s}\sigma_v^2/a\mathrm{d}v} - 1\right)\mathrm{d}u. \tag{109}$$

Finally, Definition 1 gives $\sigma_0 = a^{-1/2}(\tau_F + \phi_0)^{-1} \cdot (1 + a\phi_0')$ and using the parameter transformations in (17) gives the recovery formula in (B).

To get the bounds in (22) we note that the numerator of the ratio inside the integral in (B) is nonnegative, so a necessary condition for the expression in the integral to be nonnegative is:

$$1/c > \int_0^s \exp\left(-\int_0^v \sigma_u/a^{1/2}du\right)\sigma_v^2/adv, \quad \forall s \in [0, T] \tag{110}$$

with strict inequality for the ratio to be further finite. Since the right-hand side of the inequality is nondecreasing in s, this inequality is satisfied at all dates s if:

$$1/c > \int_0^T \exp\left(-\int_0^v \sigma_u/a^{1/2}du\right)\sigma_v^2/adv \equiv L(a). \tag{111}$$

Furthermore, for the expression inside the integral to be nonnegative at all dates we need:

$$1/c \leq \exp\left(-\int_0^s \sigma_u/a^{1/2}\,du\right)\sigma_s/a^{1/2} + \int_0^s \exp\left(-\int_0^v \sigma_u/a^{1/2}\,du\right)\sigma_v^2/a\,dv, \quad \forall s \in [0,T], \tag{112}$$

which is satisfied at all dates if:

$$1/c \leq \inf_{s\in[0,T]} \exp\left(-\int_0^s \sigma_u/a^{1/2}\,du\right)\sigma_s/a^{1/2} + \int_0^s \exp\left(-\int_0^v \sigma_u/a^{1/2}\,du\right)\sigma_v^2/a\,dv \equiv U(a). \tag{113}$$

Regarding the last part, let $\tau \in [0,T]$ be the time at which the infimum in (113) is reached and consider:

$$U(a) - L(a) = e^{-\int_0^\tau \sigma_u/a^{1/2}\mathrm{du}}\left(\sigma_\tau/a^{1/2} - \int_\tau^T e^{-\int_\tau^v \sigma_u/a^{1/2}\mathrm{du}}\sigma_v^2/a\mathrm{dv}\right). \tag{114}$$

Using that $\left(e^{-\int_\tau^v \sigma_u/a^{1/2}\mathrm{du}}\right)' = -e^{-\int_\tau^v \sigma_u/a^{1/2}\mathrm{du}}\sigma_v/a^{1/2}$ and integrating by parts this expression can be simplified to:

$$U(a) - L(a) = e^{-\int_0^T \sigma_u/a^{1/2}\mathrm{du}}/a^{1/2}\left(\sigma_T - \int_\tau^T e^{\int_v^T \sigma_u/a^{1/2}\mathrm{du}}\sigma_v'\mathrm{dv}\right). \tag{115}$$

The interval in (22) becomes empty if this expression becomes negative. Note that this expression hits zero (interval becomes a single point) when:

$$\sigma_T = \int_\tau^T e^{\int_v^T \sigma_u/a^{1/2}\mathrm{du}}\sigma_v'\mathrm{dv}, \tag{116}$$

which corresponds to (23), and that:

$$\lim_{a\to\infty} U(a) - L(a) = \sigma_\tau \geq 0. \tag{117}$$

Hence, either (23) has no solution and by the continuity of (115) the interval is always nonempty, or instead (23) has possibly multiple solutions, $\{a_k\}$, in which case by continuity of (115) the interval is nonempty on $a \geq \max_k a_k$ (and possibly other intervals).

A.3 Proof of Proposition 1

We descretize the law of motion of unobservables, $\mathbf{\Psi}_t$, over the (daily) time intervals, Δ, as:

$$\mathbf{\Psi}_{n\cdot\Delta} = \mathbf{\Psi}_{(n-1)\cdot\Delta} + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{\equiv \mathbf{B}} \epsilon_{m,n\cdot\Delta}, \quad \epsilon_{m,n\cdot\Delta} \sim \mathcal{N}(0, \tau_m^{-1} \cdot \Delta) \tag{118}$$

for $n = 1, ..., N$. Each day the empiricist observes a new realization of CAR:

$$P_{n\cdot\Delta} = \underbrace{\begin{pmatrix} 1 - \lambda_{1,n\cdot\Delta} & 0 \end{pmatrix}}_{\equiv \mathbf{G}_{n\cdot\Delta}} \widehat{\mathbf{\Psi}}^c_{n\cdot\Delta} + \underbrace{\begin{pmatrix} \lambda_{1,n\cdot\Delta} & \lambda_{2,n\cdot\Delta} \end{pmatrix}}_{\equiv \mathbf{H}_{n\cdot\Delta}} \mathbf{\Psi}_{n\cdot\Delta}, \tag{119}$$

where we keep the notation $\widehat{\mathbf{\Psi}}^c_{n\cdot\Delta} \equiv \mathbb{E}[\mathbf{\Psi}_{n\cdot\Delta}|\mathscr{F}^c_{n\cdot\Delta}]$ and $\mathbf{O}^c_{n\cdot\Delta} \equiv \mathbb{V}[\mathbf{\Psi}_{n\cdot\Delta}|\mathscr{F}^c_{n\cdot\Delta}]$, with $\mathscr{F}^c_{n\cdot\Delta} \equiv \{P_{k\cdot\Delta} : k \leq n\}$. Applying discrete-time Kalman filtering gives:

$$\begin{aligned} \widehat{\mathbf{\Psi}}^c_{n\cdot\Delta} &= \widehat{\mathbf{\Psi}}^c_{(n-1)\cdot\Delta} \\ &\quad + \underbrace{\mathrm{Cov}(\mathbf{\Psi}_{n\cdot\Delta}, P_{n\cdot\Delta}|\mathscr{F}^c_{(n-1)\cdot\Delta})\mathrm{Var}(P_{n\cdot\Delta}|\mathscr{F}^c_{(n-1)\cdot\Delta})^{-1}}_{\equiv \overline{\mathbf{K}}_{n\cdot\Delta}}(P_{n\cdot\Delta} - \mathbb{E}[P_{n\cdot\Delta}|\mathscr{F}^c_{(n-1)\cdot\Delta}]) \end{aligned} \tag{120}$$

$$\begin{aligned} \mathbf{O}^c_{n\cdot\Delta} &= \mathrm{Var}[\mathbf{\Psi}_{n\cdot\Delta}|\mathscr{F}^c_{(n-1)\cdot\Delta}] \\ &\quad - \mathrm{Cov}[\mathbf{\Psi}_{n\cdot\Delta}, P_{n\cdot\Delta}|\mathscr{F}^c_{(n-1)\cdot\Delta}]\mathrm{Var}[\mathbf{\Psi}_{n\cdot\Delta}|\mathscr{F}^c_{(n-1)\cdot\Delta}]^{-1}\mathrm{Cov}[P_{n\cdot\Delta}, \mathbf{\Psi}_{n\cdot\Delta}|\mathscr{F}^c_{(n-1)\cdot\Delta}]. \end{aligned} \tag{121}$$

Direct computations using discretized dynamics in (118) give:

$$\mathrm{Cov}[\mathbf{\Psi}_{n\cdot\Delta}, P_{n\cdot\Delta}|\mathscr{F}^c_{(n-1)\cdot\Delta}] = (1 - \mathbf{G}_{n\cdot\Delta}\overline{\mathbf{K}}_{n\cdot\Delta})^{-1}(\mathbf{O}^c_{(n-1)\cdot\Delta} + Q\mathbf{B}\mathbf{B}')\mathbf{H}'_{n\cdot\Delta} \tag{122}$$

$$\mathrm{Var}[\mathbf{\Psi}_{n\cdot\Delta}|\mathscr{F}^c_{(n-1)\cdot\Delta}] = (1 - \mathbf{G}_{n\cdot\Delta}\overline{\mathbf{K}}_{n\cdot\Delta})^{-2}(\mathbf{H}_{n\cdot\Delta}(\mathbf{O}^c_{(n-1)\cdot\Delta} + Q\mathbf{B}\mathbf{B}')\mathbf{H}'_{n\cdot\Delta}). \tag{123}$$

Plugging these expressions in the filter gain and several simplification steps give:

$$\overline{\mathbf{K}}_{n\cdot\Delta} = (1 - \mathbf{G}_{n\cdot\Delta}\overline{\mathbf{K}}_{n\cdot\Delta})^{-1}\mathbf{K}_{n\cdot\Delta}, \tag{124}$$

where $\mathbf{K}_{n\cdot\Delta}$ denotes the Kalman gain with respect to the sufficient price statistic (in the empirical implementation the empiricist observes CAR as opposed to the sufficient price statistic directly), which satisfies:

$$\mathbf{K}_{n\cdot\Delta} = (\mathbf{O}^c_{(n-1)\cdot\Delta} + Q\mathbf{B}\mathbf{B}')\mathbf{H}'_{n\cdot\Delta}(\mathbf{H}_{n\cdot\Delta}(\mathbf{O}^c_{(n-1)\cdot\Delta} + Q\mathbf{B}\mathbf{B}')\mathbf{H}'_{n\cdot\Delta})^{-1}. \tag{125}$$

From these expressions the dynamics of Kalman filter in (120) simplify to:

$$\widehat{\mathbf{\Psi}}^c_{n\cdot\Delta} = \widehat{\mathbf{\Psi}}^c_{(n-1)\cdot\Delta} + \overline{\mathbf{K}}_{n\cdot\Delta}(P_{n\cdot\Delta} - (\mathbf{G}_{n\cdot\Delta} + \mathbf{H}_{n\cdot\Delta})\widehat{\mathbf{\Psi}}^c_{(n-1)\cdot\Delta}) \tag{126}$$

$$\mathbf{O}^c_{n\cdot\Delta} = (\mathbf{I} - \mathbf{K}_{n\cdot\Delta}\mathbf{H}_{n\cdot\Delta})(\mathbf{O}^c_{(n-1)\cdot\Delta} + Q\mathbf{B}\mathbf{B}'). \tag{127}$$

We can then use observational equivalence to simplify this further to a univariate problem, using:

$$\mathbf{O}^c_{n\cdot\Delta} \equiv \begin{pmatrix} 1 & -\lambda_{1,n\cdot\Delta}/\lambda_{2,n\cdot\Delta} \\ -\lambda_{1,n\cdot\Delta}/\lambda_{2,n\cdot\Delta} & (\lambda_{1,n\cdot\Delta}/\lambda_{2,n\cdot\Delta})^2 \end{pmatrix} /\tau^c_{n\cdot\Delta}. \tag{128}$$

Substituting in the dynamics in (126) and several simplifications steps give:

$$\begin{aligned} \widehat{F}^c_{n\cdot\Delta} &= \widehat{F}^c_{(n-1)\cdot\Delta} \\ &+ \frac{\frac{\tau_{P,n\cdot\Delta}\tau_m^{1/2}}{\lambda_{2,n\cdot\Delta}\tau^c_{n\cdot\Delta}}}{1 + (1-\lambda_{1,n\cdot\Delta})\frac{\tau_{P,n\cdot\Delta}\tau_m^{1/2}}{\lambda_{2,n\cdot\Delta}\tau^c_{n\cdot\Delta}}} \left(P_{n\cdot\Delta} - \widehat{F}^c_{(n-1)\cdot\Delta} - \frac{\lambda_{1,n\cdot\Delta}}{\lambda_{2,n\cdot\Delta}}(P_{(n-1)\cdot\Delta} - \lambda_{1,(n-1)\cdot\Delta}\widehat{F}^c_{(n-1)\cdot\Delta}) \right) \end{aligned} \tag{129}$$

$$\frac{1}{\tau^c_{n\cdot\Delta}} = \frac{1}{\tau^c_{(n-1)\cdot\Delta}} + \tau^2_{P,n\cdot\Delta}\cdot\Delta, \tag{130}$$

where $\tau_{P,n\cdot\Delta}$ denotes the price signal-to-noise ratio:

$$\tau_{P,n\cdot\Delta} \equiv \tau_m^{1/2}\frac{1}{\Delta}\left(\frac{\lambda_{1,n\cdot\Delta}}{\lambda_{2,n\cdot\Delta}} - \frac{\lambda_{1,(n-1)\cdot\Delta}}{\lambda_{2,(n-1)\cdot\Delta}}\right), \tag{131}$$

the discrete-time equivalent to (69). Next we use the equilibrium solution for $\lambda_{1,n\cdot\Delta}$ and $\lambda_{2,n\cdot\Delta}$, which after several simplifications steps and substituted in the above gives:

$$\widehat{F}^c_{n\cdot\Delta} = \frac{\tau^c_{(n-1)\cdot\Delta}}{\tau^c_{n\cdot\Delta}}\widehat{F}^c_{(n-1)\cdot\Delta} + \left(1 - \frac{\tau^c_{(n-1)\cdot\Delta}}{\tau^c_{n\cdot\Delta}}\right)\frac{\tau_{n\cdot\Delta}P_{n\cdot\Delta} - \tau_{(n-1)\cdot\Delta}P_{(n-1)\cdot\Delta}}{\tau_{n\cdot\Delta} - \tau_{(n-1)\cdot\Delta}} \tag{132}$$

$$\tau^c_{n\cdot\Delta} = \tau^c_{(n-1)\cdot\Delta} + a\left(\frac{\phi_{n\cdot\Delta} - \phi_{(n-1)\cdot\Delta}}{\Delta}\right)^2\Delta. \tag{133}$$

Diffuse priors implies initial conditions $\widehat{F}_0^c = 0$ and $\tau_0^c = \tau_F$, and after iterating over the difference equations above we get:

$$\widehat{F}_{n\cdot\Delta}^c = \mathbf{1}_{n\geq 1}\frac{1}{\tau_{n\cdot\Delta}^c}\sum_{k=1}^{n}\frac{\tau_{k\cdot\Delta}^c - \tau_{(k-1)\cdot\Delta}^c}{\tau_{k\cdot\Delta} - \tau_{(k-1)\cdot\Delta}}(\tau_{k\cdot\Delta}P_{k\cdot\Delta} - \tau_{(k-1)\cdot\Delta}P_{(k-1)\cdot\Delta}) \tag{134}$$

$$\tau_{n\cdot\Delta} = \tau_F + a\cdot\sum_{k=1}^{n}\left(\frac{\phi_{k\cdot\Delta} - \phi_{(k-1)\cdot\Delta}}{\Delta}\right)^2\Delta, \tag{135}$$

which after applying the parameter transformation in (17) gives (35) and (36), with the average investors' precision following from its own definition, $\tau_{n\cdot\Delta} \equiv \tau_{n\cdot\Delta}^c + \phi_{n\cdot\Delta}$.

Based on this filter the empiricist predicts the next CAR datapoint according to:

$$\begin{aligned}
\mu_{n\cdot\Delta|(n-1)\cdot\Delta} &= \mathbb{E}[P_{n\cdot\Delta}|\mathscr{F}_{(n-1)\cdot\Delta}^c] &(136)\\
&= (\mathbf{G}_{n\cdot\Delta} + \mathbf{H}_{n\cdot\Delta})\mathbf{\Psi}_{(n-1)\cdot\Delta}^c &(137)\\
&= \left(1 - \frac{\lambda_{2,n\cdot\Delta}}{\lambda_{2,(n-1)\cdot\Delta}}\right)\widehat{F}_{(n-1)\cdot\Delta}^c + \frac{\lambda_{2,n\cdot\Delta}}{\lambda_{2,(n-1)\cdot\Delta}}P_{(n-1)\cdot\Delta}\mathbf{1}_{n\geq 1}, &(138)
\end{aligned}$$

which using equilibrium expressions for $\lambda_{1,n\cdot\Delta}$ and $\lambda_{2,n\cdot\Delta}$ gives (34). This prediction is made with error variance:

$$\begin{aligned}
\Sigma_{n\cdot\Delta|(n-1)\cdot\Delta} &= \mathrm{Var}[P_{n\cdot\Delta}|\mathscr{F}_{(n-1)\cdot\Delta}^c] &(139)\\
&= (1 + \mathbf{G}_{n\cdot\Delta}\mathbf{K}_{n\cdot\Delta})^2\mathbf{H}_{n\cdot\Delta}(\mathbf{O}_{(n-1)\cdot\Delta}^c + Q\mathbf{B}\mathbf{B}'\mathbf{1}_{n\geq 1})\mathbf{H}_{n\cdot\Delta}' &(140)\\
&= \left(1 - \tau_{P,n\cdot\Delta}a^{1/2}\right)^2\lambda_{2,n\cdot\Delta}^2\Delta/\tau_m\frac{\tau_{n\cdot\Delta}^c}{\tau_{(n-1)\cdot\Delta}^c} &(141)\\
&= \left(\frac{\tau_{n\cdot\Delta} - \tau_{(n-1)\cdot\Delta}}{\tau_{n\cdot\Delta}}\right)^2\frac{\tau_{n\cdot\Delta}^c}{(\tau_{n\cdot\Delta}^c - \tau_{(n-1)\cdot\Delta}^c)\tau_{(n-1)\cdot\Delta}^c}, &(142)
\end{aligned}$$

where the last equality follows from substituting equilibrium expressions for $\lambda_{1,n\cdot\Delta}$ and $\lambda_{2,n\cdot\Delta}$. From this expression we can make the claim preceding the proposition precise. In particular, reorganizing (142) we have:

$$\frac{\Sigma_{n\cdot\Delta|(n-1)\cdot\Delta}}{\Delta} = \left(\frac{\frac{\tau_{n\cdot\Delta} - \tau_{(n-1)\cdot\Delta}}{\Delta}}{\tau_{n\cdot\Delta}}\right)^2\frac{\tau_{n\cdot\Delta}^c}{\frac{\tau_{n\cdot\Delta}^c - \tau_{(n-1)\cdot\Delta}^c}{\Delta}\tau_{(n-1)\cdot\Delta}^c}. \tag{143}$$

Taking limits on both sides gives:

$$\lim_{\Delta \to 0} \left(\frac{\Sigma_{n \cdot \Delta | (n-1) \cdot \Delta}}{\Delta} \right)^{1/2} = \sigma_t \tag{144}$$

$$= \tau_t^{-1} a^{-1/2} (1 + a \cdot \phi_t'), \tag{145}$$

which is Definition 1 and the approximation in (31).

A.4 Descriprive Statistics

Years	Δ Market Cap	Δ log(BM)	Δ Illiq	Observations
1996-2000	-174.28***	0.18***	-2.06***	2'396
	(-2.63)	(7.38)	(-18.26)	
2001-2005	-553.82***	0.07***	-2.73***	2'269
	(-5.84)	(2.69)	(-21.71)	
2006-2010	-622.17***	0.04	-1.27***	2'139
	(-4.51)	(1.43)	(-6.97)	
2011-2016	-967.62***	-0.02	-1.29***	1'817
	(-4.07)	(-0.88)	(-7.49)	
2017-2022	-1'468.47***	0.03	-0.33***	1'857
	(-4.58)	(1.01)	(-3.30)	

Table 5: Summary statistics. This table shows the average difference in market capitalization, log of book-to-market ratio and illiquidity estimated by **?** between the sample of shocked firms and the rest of the CRSP stock universe. Market capitalization and book-to-market ratios are calculated at the end of each year and lliquidity is calculated annually using daily data. A stock that is shocked in a given year belongs the the shock sample in this specific year, but is not included again in the following years, unless it is shocked again. We calculate t-statistics using two-sample t-tests over five-year bins which are shown in parentheses. We follow Levene (1960) to test whether the variances in the two samples are equal and accordingly adjust the assumption of (un)equal variances for the t-test. *,**,*** denote 10%,5% and 1% significance levels.