Business, Liquidity and Information Cycles

Gorkem Bostanci* Guillermo Ordoñez†

March 24, 2024

Abstract

Stock markets play a dual role: provide information about firms' fundamentals, which is useful to allocate capital, and facilitate access to liquidity. We propose a trading model in which changes in fundamentals and/or liquidity endogenously affect both the information and the noise about fundamentals contained in prices. We structurally estimate the information content in stock prices for several countries and show that it declines in periods of liquidity distress. We incorporate this module into a dynamic general equilibrium model with heterogeneous firms to study the real effects of stock price informativeness through capital misallocation across firms. Calibrating the model for the U.S., we show that less informative stock markets induced by liquidity distress magnify the declines in output experienced during recessions by 43%.

*University of British Columbia, Email: gorkem.bostanci@ubc.ca
†University of Pennsylvania and NBER, Email: ordonez@econ.upenn.edu

1 Introduction

Stock markets have been widely recognized for their critical role in the smooth functioning of market economies by i) allocating resources and ii) providing liquidity. The first comes from the remarkable ability of markets to aggregate information about firms' profitability and prospects that is dispersed among traders, usually referred to as *stock price informativeness*. The second comes from the easiness for agents to buy or sell stocks to satisfy their *liquidity needs*, which depends on the operation of market makers and brokers and the depth of a market for a given stock.

Despite a rich literature discussing these allocation and liquidity roles of stock markets separately, how they interact through the extent of price informativeness is less understood. Does price informativeness vary with the business cycle? If so, does price informativeness respond differently to downturns when accompanied by liquidity distress in markets given by heightened liquidity needs? How do stock markets guide the allocation of resources during the different types of downturns? How do they shape recoveries?

In this paper we make progress on three fronts. One is *theoretical*, by building a model of asset trading and information acquisition where both the information and the noise content of asset prices are determined endogenously to changes in firms' fundamentals and markets' liquidity. We further include this module into a dynamic general equilibrium setting with heterogeneous firms in which stock price informativeness has real allocation and investment implications. The other front is *empirical*, by structurally estimating stock price informativeness from firm-level panel data for several countries. The last front of progress is *quantitative*, by calibrating the general equilibrium model to quantify the role of price informativeness on capital allocation in recessions with and without financial distress that heighten the use of stocks to fulfill liquidity needs.

To be more precise, our theoretical contribution consists on extending the workhorse model of price informativeness by Grossman and Stiglitz (1980). In their work, traders acquire information about a firm's fundamentals and stock price informativeness is determined by the presence and extent of exogenous noise traders. Instead, to accommodate the two main roles of stock markets, we allow for two types of traders -day and night- interested in different asset properties -liquidity and fundamentals respectively. This structure creates endogenous noise in prices: a high price may indicate that the stock can be easily traded or that the firm has strong fundamentals. The trades from one type of trader act as noise for the other. In the model, price informativeness is endogenous as both types of traders choose whether to acquire information, at a cost, to reduce their uncertainty relative to only relying on observing prices. We show a linear pricing function exists where the relative weights of information on the asset's payoff and liquidity determine price informativeness and are given by the relative fractions of day and night traders who are informed, and how aggressively they trade on their information.

We then incorporate this trading module into a real business cycle model with heterogeneous firms. We set up a link between financial and real markets where the linear pricing function is preserved in a non-linear production economy. Stocks are claims to firms' earnings, which depend on idiosyncratic firms' productivities (i.e., firms' fundamentals). If those productivities were known, investors would allocate capital across firms efficiently. In our model, those productivities are ex-ante unknown, and allocations are based on the investors' best guesses based on stock prices, which not only contain information about firms' productivities acquired by night traders but also about stocks' liquidity acquired by day traders. Hence, liquidity concerns in the economy affect resource allocations and real activity through their effects on price informativeness.

Since our model provides a linear relationship between stock prices, firms' earnings, and stock liquidity, we advance on the empirical front by showing that price informativeness can be estimated outside the general equilibrium model every year in each economy, using cross-sectional information on stock prices, earnings expectations, and stock price volatilities. We implement this structural estimation and obtain yearly stock price informativeness for 22 countries, and assess its cyclical properties. We show that stock price informativeness over the business cycles is heavily determined by the extent of liquidity distress: in periods of high liquidity needs, such as the Great Recession and COVID-19 pandemic episodes, price informativeness substantially declines.

Finally, on our quantitative contribution, we measure the relevance of stock price informativeness and its dependence on financial distress on the allocation of resources, total investment and other macro aggregates. We calibrate the parameters of the full model to the United States, assuming the economy is subject to two, possibly correlated, aggregate shocks: firms' productivity and liquidity needs. This is meant to capture recessions with and without distress in financial markets. We use aggregate moments together with moments of the estimated pricing functions and price informativeness to jointly identify periods with low productivity and/or high liquidity needs. We can use these periods to jointly discipline the cost of acquiring information and the series of liquidity needs measured by the fraction of day traders operating in stock markets.

Using the calibration, we conduct impulse-response exercises to simulate downturns, with and without financial distress. We show that during recessions with financial distress stock markets become less informative about firms' fundamentals, inducing a worse allocation of resources, which also discourages investment. In other words, a larger fraction of traders in search of liquidity participating in stock markets absconds information about firms' productivity, in spite of all traders acquiring more information. In contrast, if the recession is not accompanied by financial distress, in our calibration, stock markets become more informative, and all traders acquire more information. This result shows that downturns are not "cleansing recessions" if accompanied by financial distress and pressures from liquidity needs. In quantitative terms, our exercise shows that an average recession with financial distress would result in an output decline that is 40% larger and an investment decline that is 50% larger than the same recession without financial distress. This is, we believe, a sizeable real effect of changes in liquidity needs, which have an impact through changes in price informativeness and misallocation.

Based on our calibration we have also explored how the economy would fare when facing a recession with financial distress under alternative information structures. If information were exogenous, the change in the composition of traders and their trading intensities would reduce price informativeness more drastically, leading to more severe misallocations and investment drops, with output declining 30% more than our benchmark with endogenous information. This is because information acquisition acts as a stabilizer against shocks that tend to reduce price informativeness. If the cost of information *about a firm's fundamentals* declines by half, price informativeness would decline less, leading to output declines that are 15% smaller. If the cost of information *about a stock's liquidity* declines by half, price informativeness would decline more, leading to output declines that are 15% larger. These results suggest that regulations facilitating information about firms' profitability make the economy more resilient to recessions with financial shocks, but those facilitating information about stock markets' depth and volume do the opposite.

Literature Review: Our paper lies at the intersection of the literature on price informativeness and the literature on input misallocation. The vast majority of the theoretical literature on price informativeness assumes an exogenous source of noise to prevent prices from being perfectly informative, following the impossibility theorem of Grossman and Stiglitz (1980). We endogenize the information/noise ratio by assuming two dimensions of information condensed in a single price. The closest papers to ours here are Stein (1987) and Vives (2014). Both use heterogeneity in traders' characteristics (the former on market access and the latter on preferences) to generate imperfectly informative prices without exogenous noise. Instead, we additionally allow for heterogeneity in asset dimensions and explore the implications of their change over time.

The empirical literature focuses on measuring price informativeness. Dávila and Parlatore (2018) use time-series regressions to measure price informativeness for each stock, which requires them to make assumptions on how model parameters change over time to keep the cross-sectional variation flexible. We, on the other hand, use cross-sectional regressions to measure price informativeness of the stock market over time, which requires us to make assumptions on the extent of heterogeneity across stocks to allow parameters to change flexibly over time. Bai et al. (2016), similar to us, analyzes the long-run trend in price informativeness using cross-sectional regressions. However, they are interested in the ability of prices to predict future stock performance, which is determined jointly by the availability of information on future prices and the ability of stock markets to communicate such information. Our model allows disentangling the two components.

We also contribute to the literature on input misallocation, which analyzes patterns across firms in recessions. Foster et al. (2016) find the extent of labor reallocation across the U.S. firms has declined during the Great Recession. Eisfeldt and Rampini (2006) and Cooper and Schott (2023) further show the amount of capital reallocation is procyclical. Kehrig (2015) finds dispersion of productivity distribution in the U.S. is larger in recessions than booms.[1] Tighter financial constraints, counter-cyclical adverse selection in the market for used-capital, managers' incentives to hide reallocation needs from owners during recessions have been proposed as potential mechanisms for the counter-cyclical misallocation.[2] On the other hand, others take increased uncertainty/misallocation as a

[1] The increased misallocation is not specific to the U.S. and has been documented for other countries in economic crises as well. See Oberfield (2013) for Chile, Sandleris and Wright (2014) for Argentine, Dias et al. (2016) for Portugal and Di Nola (2016) for the U.S.

[2] See Ordonez (2013), Khan and Thomas (2013), Fajgelbaum et al. (2017) and Straub and Ulbricht (2023)

primitive shock and analyze its effects to understand business cycles.[3] We contribute to this literature by showing that these results should be qualified by the presence of financial crises during recessions, by providing a novel mechanism that affects misallocation and by quantitatively assessing its importance.

There is a recent, but growing, literature at the intersection of these two strands. Benhabib et al. (2019) does a theoretical analysis with two-way learning between the real and the financial sectors. Their model exhibits complementarity in information acquisition. Thus, a shock that reduces incentives to acquire information in one sector induces the other to reduce information acquisition, enabling equilibrium switches that amplify the initial shock. David et al. (2016) and David and Venkateswaran (2019) are the closest papers to our study in this intersection. The former focuses on the role of informational frictions in resource allocation and measures how much each source of information contributes to productivity gaps. The latter has a larger scope and incorporates many potential frictions that can distort resource allocation on top of informational frictions. Both analyses provide static measures; thus, they are silent about cyclicality. While our framework restricts attention to stock markets as the main source of information, we introduce endogenous noise and time-varying model parameters.

The remainder of the paper is structured as follows. Section 2 introduces the stock market model with endogenous noise, and incorporates it into an otherwise standard RBC model. Section 3 describes the data sources, the empirical strategy to compute stock price informativeness, and study its cyclical properties and relation with measures of financial distress. Section 4 calibrates the model to the United States and Section 5 uses these parameters to assess the quantitative relevance of financial distress and information costs on real variables through their impact on price informativeness. Section 6 concludes.

2 Model

In this section, we build a model of stock trading and incorporate it into a real business cycle model with firm heterogeneity. A firm's stock price provides information on two

for tighter financial constraints, Fuchs et al. (2016) for adverse selection, and Eisfeldt and Rampini (2008) for managerial incentives.

[3] See Christiano et al. (2014), Arellano et al. (2019), and Bloom et al. (2018).

dimensions: the firm's long-term profitability and the stock's short-term liquidity, which are valued differently by different agents. The trading activity of one type of agent masks the information for the other type and affects information acquisition choices by both types. Since the real sector uses the stock price as a signal for a firm's long-term profitability, the degree of misallocation (of inputs) in the economy depends on trading needs and information production in stock markets. To keep the notation simple, we suppress time subscripts unless necessary.

2.1 Environment

Preferences There is a measure one of traders who live one period and a measure one of infinitely lived households.

Traders have constant-absolute-risk-aversion (CARA) preferences where the utility from consuming an amount W is given by $\nu(W) = -e^{-aW}$ for $a > 0$. At the start of each period, newborn traders receive a liquidity shock: a fraction $\gamma \in (0, 1)$ of them ('day traders') need to consume early while the rest ('night traders') need to consume at the end of the day.

The representative household has constant-relative-risk-aversion (CRRA) preferences with inter-temporal elasticity of substitution $1/\eta$; that is, utility from consuming an amount W is given by $u(W) = \frac{W^{1-\eta}-1}{1-\eta}$.

Technology There is a measure one of firms (indexed by i) with profit function:

$$\Pi_i = z_{in}(\bar{K}_i + K_i) - r_i K_i - \xi K_i^2 / 2\bar{K}_i \tag{1}$$

where z_{in} is the productivity of firm i, $\bar{K}_i$ is installed capital, and K_i is rented capital at a price r_i. While $\bar{K}_i$ cannot be changed or reallocated, K_i is determined period-by-period.

Assets, Endowments, and Market Structure There are three types of assets in the economy: capital, firms' shares, and a foreign bond.

Capital is owned by households and rented to a mutual fund for a fixed rental rate of r. The mutual fund allocates the capital across firms to maximize total returns. We assume the mutual fund does not hold market power against households (we can think of households as owners of the mutual fund); the interest rate r distributes all the surplus to the household. On the other hand, the mutual fund is a price-discriminating monopoly when transacting with the firms. In particular, to each firm i, the mutual fund can make a take-it-or-leave-it offer $\{K_i, r_i(z_{in})\}$.[4]

Traders have access to two assets: a foreign bond supplied at a fixed return r^F and firms' shares, i.e., stocks. Each share gives ownership of 1 unit of installed capital at the associated firm, making the outstanding share amount of firm i equal to $\bar{K}_i$.[5] Each share i is subject to an exogenous liquidity discount, which is represented by a decline in the return by z_{id} if sold prematurely.

Information The profitability of the firm (z_{in}) consists of three parts: the aggregate productivity shock, which is public information (Z), a random term that can be learned (θ_{in}), and a random term that cannot be learned ($\tilde{\varepsilon}_{in}$). The aggregate productivity Z and the fraction of day traders γ define a state $s = \{Z, \gamma\}$ and follow a joint Markov process with transition probability $q_{s,s'}$. The learnable component θ_{in} is drawn from a prior distribution $\mathcal{N}(\bar{\theta}_{in}, \sigma^2_{\theta_{in}})$, while the unlearnable component $\tilde{\varepsilon}_{in}$ follows an AR(1) process: $\tilde{\varepsilon}_{in} = \rho\tilde{\varepsilon}^-_{in} + \varepsilon_{in}$ with $\varepsilon_{in} \sim \mathcal{N}(0, \sigma^2_{\varepsilon_{in}})$, where $\tilde{\varepsilon}^-_{in}$ is public information.[6]

The liquidity discount of the stock (z_{id}) consists of two parts: a random term that can be learned (θ_{id}), and a random term that cannot be learned ($\tilde{\varepsilon}_{id}$). The first component is drawn from a prior distribution $\mathcal{N}(\bar{\theta}_{id}, \sigma^2_{\theta_{id}})$ and the second from a prior distribution $\mathcal{N}(0, \sigma^2_{\varepsilon_{id}})$.

[4] Note that while K_i cannot be conditioned on z_{in}, the interest rate can be. This market structure guarantees that the profitability of holding a firm's stock is independent of the capital allocated to such a firm, which is necessary for the existence of a linear pricing function.

[5] The allocation of stocks across traders is irrelevant for aggregate quantities since CARA utility functions give rise to policy functions that are independent of wealth.

[6] We capture persistent differences in idiosyncratic returns across stocks through $\tilde{\varepsilon}_{in}$, the component that cannot be learned. This is in contrast to Farboodi and Veldkamp (2020), who focus on the learnable component to capture such dynamics. Unlike their setting, the future pricing function in our setting is stochastic, and the pricing function becomes nonlinear if there is any persistence in the learnable component.

To summarize the distributions of the profitability and liquidity of a firm,

$$
\begin{aligned}
z_{in} =& Z + \theta_{in} + \tilde{\varepsilon}_{in} \\
& where \quad \theta_{in} \sim \mathcal{N}(\bar{\theta}_{in}, \sigma^2_{\theta_{in}}) \qquad \tilde{\varepsilon}_{in} = \rho \tilde{\varepsilon}^-_{in} + \varepsilon_{in} \quad \& \quad \varepsilon_{in} \sim \mathcal{N}(0, \sigma^2_{\varepsilon_{in}}) \\
z_{id} =& \theta_{id} + \varepsilon_{id} \\
& where \quad \theta_{id} \sim \mathcal{N}(\bar{\theta}_{id}, \sigma^2_{\theta_{id}}) \qquad \varepsilon_{id} \sim \mathcal{N}(0, \sigma^2_{\varepsilon_{id}})
\end{aligned}
\tag{2}
$$

We assume $\theta_{id}, \theta_{in}, \varepsilon_{id}, \varepsilon_{in}$ are iid across firms and independently distributed over time. We allow $\sigma^2_{\theta_n}$, $\sigma^2_{\theta_d}$, $\sigma^2_{\varepsilon_n}$, and $\sigma^2_{\varepsilon_d}$ to be functions of aggregate productivity Z.[7]

For each firm i, night traders can pay $c(\lambda_{in})$ and day traders can pay $c(\lambda_{id})$ to learn $\theta = \{\theta_{id}, \theta_{in}\}$ while ε_{id} and ε_{in} are learned for free only after their realization. We denote with λ_{il} the fraction of $l \in \{d, n\}$ traders who choose to be informed about stock i and assume $c(\lambda_{il})$ is increasing in λ_{il}.[8] Finally, we assume the mutual fund doesn't have access to this information technology, and similar to uninformed traders, it infers z_{in} by observing stock market prices.

Timing Each period starts with $\theta = \{\theta_{in}, \theta_{id}\}^1_{i=0}$ realized. Then, traders simultaneously acquire information and trade stocks. The mutual fund allocates capital across firms after observing the stock prices. Then, $\varepsilon = \{\varepsilon_{in}, \varepsilon_{id}\}^1_{i=0}$ is realized, with both θ and ε becoming public information. The new generation of traders is born; the day traders sell their stocks to the newborn traders at a discount, consume, and die. The production takes place, the firms pay the mutual fund, and the fund pays households. The night traders sell their stocks to the newborn traders, consume, and die.

[7] We make these volatilities only a function of productivity Z and not of liquidity needs γ for two reasons. First, liquidity needs are modeled as shocks on preferences that should not affect fundamentals. Second, this choice allows later for a cleaner comparison of the role of shocks on liquidity needs on stock price informativeness and the extent of misallocations.

[8] The cost function increasing in the fraction of informed traders rules out possible equilibrium multiplicity that may arise from complementarities in information acquisition. It can be motivated through heterogeneous information acquisition costs, with the traders with the lowest cost being the ones who acquire the information first. An interior equilibrium exists if the marginal trader is indifferent between acquiring costly information and relying on prices.

2.2 Agents' Problems and Market Clearing

Mutual Fund's Capital Allocation Across Firms Capital is allocated across firms to maximize total expected returns. To be more precise, the mutual fund calculates the constrained efficient K_i for firm i according to

$$K_i = \max\left\{\bar{K}_i\left(\frac{E[z_{in}|p]-r}{\xi}\right), 0\right\} \quad \forall i, \tag{3}$$

then makes a take-it-or-leave-it offer with a contract $\{K_i, r_i(z_{in})\}$ that satisfies

$$z_{in}K_i - r_i(z_{in})K_i - \xi K_i^2/2\bar{K}_i = 0 \;\; \forall i. \tag{4}$$

Hence, the mutual fund obtains all the ex-post surplus from the rented capital, and the profit of firm i becomes $z_{in}\bar{K}_{in}$. Night traders receive $z_{in} + p_i'$ for each share of firm i they hold. The first term represents the dividend, and the second term represents the resale price. Day traders, due to the premature selling of the stock before production takes place, receive $z_{in} - z_{id} + p_i'$.

Traders' Portfolio Choice Problem Traders choose whether to acquire information and how much to demand of each asset, conditional on information or lack thereof. The portfolio problem of a night trader who is endowed with $\tilde{b}$ international bonds and $\tilde{X}_i$ stocks is

$$\begin{aligned} &\max_{B_n,\{X_{in}\}_{i\in[0,1]}} E\Big[-exp\Big[-a[(1+r^F)B_n + \int_i X_{in}\big(z_{in}+p_i'-p_i\big)di]\Big]\Big] \\ &s.t. \;\; B_n + \int_i p_i X_{in} di = \tilde{b} + \int_i p_i \tilde{X}_i di, \end{aligned} \tag{5}$$

and the portfolio problem for a day trader is

$$\begin{aligned} &\max_{B_d,\{X_{id}\}_{i\in[0,1]}} E\Big[-exp\Big[-a[(1+r^F)B_d + \int_i X_{id}\big(z_{in}-z_{id}+p_i'-p_i\big)di]\Big]\Big] \\ &s.t. \;\; B_d + \int_i p_i X_{id} di = \tilde{b} + \int_i p_i \tilde{X}_i di \end{aligned} \tag{6}$$

where B_l and $\{X_{il}\}_{i\in[0,1]}$ are bond and stock demands of the traders for $l \in \{d, n\}$, and r^F is an international interest rate that only traders can obtain. The information acquisition determines how the expectations are formed.

Given the distributional assumptions and the information structure, the demand for risky assets by informed and uninformed traders is given by,

$$
\begin{aligned}
X_{in}^{I*} &= \frac{E[z_{in} + p_i'|\theta] - (1+r^F)p_i}{aVar[z_{in} + p_i'|\theta]} \qquad X_{id}^{I*} = \frac{E[z_{in} - z_{id} + p_i'|\theta] - (1+r^F)p_i}{aVar[z_{in} - z_{id} + p_i'|\theta]} \\
X_{in}^{U*} &= \frac{E[z_{in} + p_i'|p_i] - (1+r^F)p_i}{aVar[z_{in} + p_i'|p_i]}, \qquad X_{id}^{U*} = \frac{E[z_{in} - z_{id} + p_i'|p_i] - (1+r^F)p_i}{aVar[z_{in} - z_{id} + p_i'|p_i]},
\end{aligned} \tag{7}
$$

where $X_{in}^{I*}, X_{id}^{I*}, X_{in}^{U*}, and X_{id}^{U*}$ represent the asset demand by night and day traders that are informed (I) and uninformed (U), respectively,

Traders' Information Acquisition Problem Let's denote the information acquisition decision of trader j for asset i with $I_{ji} \in \{0, 1\}$. Then, the end-of-period wealth for a night trader j as $W_{nj}(I_j)$. That is,

$$
W_{nj}(I_j) = (1+r^F)\left(W_{0j} - \int_i I_{ji} c(\lambda_{in}) di\right) + \int_i \left(z_{in} - (1+r^F)p_i\right)\left(I_{ji}X_{in}^{I*} + (1-I_{ji})X_{in}^{U*}\right) di \tag{8}
$$

Let I_j^1 be an arbitrary information acquisition vector where $I_{ji}^1 = 0$ for a specific stock i and I_j^2 be an identical vector, except that $I_{ji}^2 = 1$. The trader would choose I_j^2 over I_j^1, i.e., acquire information about stock i, if and only if $E\left[V\left(W_{nj}(I_j^2)\right) \mid p\right] \geq E\left[V\left(W_{nj}(I_j^1)\right) \mid p\right]$ where p is the vector of stock prices. The same comparison holds for day traders.

Lemma 1. *Let I_j^1 be an arbitrary information acquisition vector where $I_{ji}^1 = 0$ for a specific stock i and I_j^2 be an identical vector, except that $I_{ji}^2 = 1$. For night traders,*

$$
\frac{E\left[V\left(W_{nj}(I_j^2)\right) \mid p\right]}{E\left[V\left(W_{nj}(I_j^1)\right) \mid p\right]} = e^{ac(\lambda_{in})}\sqrt{\frac{Var[z_{in} + p_i'|\theta_i]}{Var[z_{in} + p_i'|p_i]}}, \tag{9}
$$

and for day traders,

$$
\frac{E\left[V\left(W_{dj}(I_j^2)\right) \mid p\right]}{E\left[V\left(W_{dj}(I_j^1)\right) \mid p\right]} = e^{ac(\lambda_{id})}\sqrt{\frac{Var[z_{in} - z_{id} + p_i'|\theta_i]}{Var[z_{in} - z_{id} + p_i'|p_i]}}. \tag{10}
$$

Proof. All proofs are in Appendix A. □

Lemma 1 shows that each trader decides whether to acquire information about each stock i, by comparing the cost of acquiring information with the decrease in the variance of their end-of-period wealth. Define $\psi^{il}(\lambda)$ as

$$\psi^{in}(\lambda) \equiv e^{ac(\lambda_{in})}\sqrt{\frac{Var[z_{in}+p_i'|\theta_i]}{Var[z_{in}+p_i'|p_i]}}, \quad \text{and} \quad \psi^{id}(\lambda) \equiv e^{ac(\lambda_{id})}\sqrt{\frac{Var[z_{in}-z_{id}+p_i'|\theta_{in},\theta_{id}]}{Var[z_{in}-z_{id}+p_i'|p_i]}}.$$

Corollary 1. *$\psi^{il}(\lambda)$ is monotone in λ_{il} for $l \in \{d, n\}$. Therefore,*
(i) If $\psi^{il}(\lambda) > 1 \ \forall \lambda_{il} \in [0,1]$, all l traders become informed, i.e. $\lambda^*_{il} = 1$.
(ii) If $\psi^{il}(\lambda) < 1 \ \forall \lambda_{il} \in [0,1]$, no l traders become informed, i.e. $\lambda^*_{il} = 0$.
(iii) Otherwise, λ^*_{il} is given by $\psi^{il}(\lambda^*_{il}) = 1$.

Problem of the Household The recursive formulation of the representative household's problem is

$$\begin{aligned} H(s,K,k) = \max_{k'} \quad & u\left(k(1+r(s,k)) - k' + \int_i \left((z_{in}-r)K_i - \frac{\xi K_i^2}{2\bar{K}_i}\right) di\right) \\ & + \beta \sum_{s'} q_{ss'} H(s',K',k') \\ & s.t. \quad K' = G(K) \end{aligned} \tag{11}$$

where $s = \{\gamma, Z\}$ denotes the aggregate state, $H(.)$ is the value function, β is the discount factor, k is the individual capital holdings and K is the aggregate capital holdings. $G(.)$ represents the household expectations over the future path of the aggregate capital. The household has the usual Euler Equation:

$$u'(k(1+r)-k') = \beta u'(k'(1+r')-k''). \tag{12}$$

Market Clearing Market clearing condition for the shares of firm i is

$$\gamma\Big[\lambda_{id}X^I_{id} + (1-\lambda_{id})X^U_{id}\Big] + (1-\gamma)\Big[\lambda_{in}X^I_{in} + (1-\lambda_{in})X^U_{in}\Big] = \bar{K}_i \tag{13}$$

The capital market clearing condition for non-installed capital is

$$\int_i K_i = K, \tag{14}$$

where K is the total capital supplied by households. The mutual fund pays to households r to break even,

$$\sum_i r_i K_i = rK. \tag{15}$$

2.3 Equilibrium

Definition Allocations and prices $H, r, k', G, \{K_i, r_i, X_{id}^I, X_{id}^U, X_{in}^I, X_{in}^U, \lambda_{id}, \lambda_{in}, \phi_{i0}, \phi_{i\varepsilon}, \phi_{id}, \phi_{in}, p_i\}_{i \in (0,1)}$ are functions of the aggregate state s and constitute a Linear Rational Expectations Equilibrium such that

1. $X_{id}^I, X_{id}^U, X_{in}^I$, and X_{in}^U solve the traders' problems, as characterized in (7).
2. $\lambda_{id}, \lambda_{in}$ are given by Corollary 1.
3. Price is a linear function of θ_{id} and θ_{in}, i.e., $p_i = \phi_{i0} + \phi_{i\varepsilon}\varepsilon_{in}^{-} + \phi_{id}\theta_{id} + \phi_{in}\theta_{in}$, where $\phi_{i0}, \phi_{i\varepsilon}, \phi_{id}$, and ϕ_{in} solve the stock market clearing condition in (13).
4. r, K_i, r_i solve the mutual fund's problem in (3) and (4) and satisfy the capital market clearing condition in (14).
5. k' and H solve the representative consumer's problem in (11) and (12).
6. G is consistent with k'.

2.4 Characterizing the Equilibrium

2.4.1 The Pricing Function

Proposition 1 shows the existence of a linear rational expectations equilibrium.

Proposition 1. *There exists a market price for stock i with the form $p_i = \phi_{i0} + \phi_{i\varepsilon}\varepsilon_{in}^{-} + \phi_{id}\theta_{id} + \phi_{in}\theta_{in}$ where*

$$\frac{|\phi_{in}|}{|\phi_{id}|} = 1 + \frac{(1-\gamma)\lambda_{in}(\sigma_{\varepsilon_{id}}^2 + Var(z_{in} + p_{in}'))}{\gamma\lambda_{id}Var(z_{in} + p_{in}')}. \tag{16}$$

The ratio $\frac{|\phi_{in}|}{|\phi_{id}|}$ suggests the relative impact of θ_{in} on the price relative to θ_{id}. Therefore, Proposition 1 provides a simple equation that describes how much can be learned about firms' fundamentals θ_{in}. Corollary 2 shows that this magnitude is determined by the extent of informed trading done by night versus day traders.

Corollary 2. *For given λ_{il}, price becomes less informative about θ_{in} when*
(i) a larger fraction of traders are day traders,
(ii) a larger fraction of day traders are informed compared to night traders,
(iii) or the day payoff is less volatile conditional on θ_{id} compared to the night payoff conditional on θ_{in}.

Corollary 3. *Let $\lim_{\lambda_{il}\to 0} c(\lambda_{il}) = 0$ and $\lim_{\lambda_{il}\to 1} c(\lambda_{il}) = \bar{C}$ where $\bar{C}$ is large enough. Then, any linear REE is interior, i.e., $\lambda_{il} \in (0,1)$.*

2.4.2 The Extent of Capital Misallocation

We define here a measure of capital misallocation and characterize it. The misallocation follows from having to use $E[\theta_{in}|p_i]$ instead of θ_{in} in allocating capital. The expression for the output loss due to misallocation is complicated, but it is monotonic in the Bayesian risk associated with using $E[\theta_{in}|p_i]$ as an estimator for θ_{in}. We first introduce a simplification that we maintain in what follows.

Assumption 1. *The parameters $\bar{K}_i, \bar{\theta}_{in}, \bar{\theta}_{id}, \sigma^2_{\varepsilon_{in}}$, $\sigma^2_{\varepsilon_{id}}$, $\sigma^2_{\theta in}$, and $\sigma^2_{\theta id}$ are firm invariant.*

Assumption 1 guarantees equilibrium fractions of informed investors λ_{in}, λ_{id} and pricing function parameters ϕ_{i0}, $\phi_{i\varepsilon}$, ϕ_{in}, ϕ_{id} are also firm invariant. Hence, it allows the use of cross-sectional variation of prices in estimating price informativeness.

We treat the mutual fund's problem as one of estimating θ_{in} with $E[\theta_{in}|p_i]$. Under a quadratic loss function, the frequentist risk would equal the summation of a squared bias ($E[E[\theta_{in}|p_i] - \theta_{in}]^2$) and a variance ($\mathrm{Var}[E[\theta_{in}|p_i] - \theta_{in}]$) term. We next derive ex-ante and interim (conditional on p_i) risk measures for the mutual fund's estimator as well as the ex-post estimation error in Proposition 2.

Proposition 2. *Under Assumption 1 and the squared loss function, the mutual fund's estimator is unbiased and the ex-ante and interim (conditional on p_i) risk involved with the inference equals*

$$R(\theta_{in}, E[\theta_{in}|p_i]) = \frac{1}{\frac{1}{\sigma^2_{\theta_n}} + \frac{1}{\sigma^2_{\theta_d}}\left(\frac{\phi_n}{\phi_d}\right)^2}.$$

The estimation error conditional on θ_{in} and θ_{id} is given by

$$E[\theta_{in}|p_i] - \theta_{in} = \frac{\frac{(\bar{\theta}_{in}-\theta_{in})}{\sigma^2_{\theta_n}} + \frac{\phi_n}{\phi_d}\frac{(\theta_{id}-\bar{\theta}_{id})}{\sigma^2_{\theta_d}}}{\frac{1}{\sigma^2_{\theta_n}} + \frac{1}{\sigma^2_{\theta_d}}\left(\frac{\phi_n}{\phi_d}\right)^2}$$

As shown in Proposition 2, the frequentist risk is independent of θ_{in}, hence equal to the Bayesian risk.[9] Corollary 4 summarizes what Proposition 2 implies about the characteristics of firms that get under and over-invested.

Corollary 4. *Under Assumption 1 and the squared loss function,*

1. *the likelihood of allocating capital to an inefficient firm increases as the signal about its stock's liquidity is more encouraging than expected, and*
2. *the likelihood of not allocating capital to an efficient firm increases as the signal about its stock's liquidity is more discouraging than expected.*

Following Goldstein et al. (2014), we define price informativeness as the reduction in the conditional variance of firm profitability after observing the price. We further normalize it with the reduction in the conditional variance of payoffs after acquiring a costly signal. We introduce this normalization to help simplify later comparisons over time and across countries.

Definition The price informativeness (PI) measure for firm i is defined as

$$PI_{in} = \frac{Var[z_{in}] - Var[z_{in}|p_i]}{Var[z_{in}] - Var[z_{in}|\theta_{in}]},$$

where $Var[z_{in}]$ denotes the unconditional variance of z_{in}.

A PI measure close to 0 would indicate that the variance reduction from observing the price is negligible compared to the reduction from acquiring the costly information. A PI measure close to 1 would suggest that observing the price is almost as useful as acquiring the costly signal. The PI measure in our setting boils down to a simple firm-invariant expression that resembles the risk function in Proposition 2.

[9] The Bayesian risk is defined as $\int R\left(\theta_{in}, E[\theta_{in}|p_i]\right) dF(\theta_{in})$.

Corollary 5. *Under Assumption 1, the PI measure equals*

$$PI_n = 1 - \frac{R(\theta_{in}, E[\theta_{in}|p_i])}{\sigma^2_{\theta_n}} = \frac{1}{1 + \frac{\sigma^2_{\theta_d}}{\sigma^2_{\theta_n}}\left(\frac{\phi_d}{\phi_n}\right)^2}. \tag{17}$$

The PI measure is inversely related to the mutual fund's risk. Hence, the extent of misallocation is decreasing in the PI measure. Under Assumption 1, the PI measure can be summarized by two parameters ($\sigma^2_{\theta_n}$ and $\sigma^2_{\theta_d}$) and two equilibrium objects (ϕ_n and ϕ_d). Our empirical strategy next consists of estimating these four objects to calculate price informativeness in a country each year.

2.5 Discussion of the Model Elements

Static Components The model replicates a standard real business cycle model with heterogeneous firms. The main difference is in the allocation problem of the mutual fund that depends on imperfect endogenous information filtering through prices. In an environment where θ_{in} is observable, the fund allocates capital across firms based on $E[z_{in}|\theta_{in}]$. Hence, 'misallocation' would be due to ε_{in}, which is inevitable. In our setting, the mutual fund can only rely on stock prices p_i to infer θ_{in} and allocate capital across firms. To allocate capital, θ_{id} is irrelevant, yet, a high stock price could stem from a high θ_{in} or a low θ_{id}. In other words, stocks may be priced higher due to higher long-term value or lower short-term fluctuations in resale value. Thus, compared to the benchmark, firms with lower (higher) than expected θ_{id} shocks are allocated more (less) capital. In summary, the existence of day traders prevents prices from perfectly revealing θ_{in}. Our main insight is that liquidity problems have real implications for garbling information and inducing capital misallocation.

Our structure allows for endogenous noise on stock market prices. In contrast to most literature, this noise does not arise from exogenous noise traders but by endogenous information acquisition by traders that have liquidity needs. This is relevant to put structure on the potential determinants of stock price informativeness and assess its real economic consequences.

Other papers relax the CARA-Normal structure to incorporate richer dynamics, but they either make assumptions on noise trader behavior and dynamics (Barlevy and Veronesi (2000), Vanden (2008), Hassan and Mertens (2017)), restrict attention to existence and uniqueness results without deriving an estimable pricing equation (Breon-Drish (2015)) or assume information acquisition is exogenous (Wang (1993), Sockin (2015)). Given our goal of estimating the role of information conveyed on stock markets for allocations, we instead eliminate the need for noise traders, have noise to vary with endogenous information, and derive a pricing equation that we can take to the data.

Aggregate Fluctuations We conceptualize the economy with an aggregate state with two components: the aggregate productivity shock Z and the fraction of day traders γ. These two components follow a joint Markov process. Furthermore, we allow the parameters related to the quality of information and firm/stock performance ($\sigma^2_{\theta_n}$, $\sigma^2_{\theta_d}$, $\sigma^2_{\varepsilon_n}$, and $\sigma^2_{\varepsilon_d}$) to change with Z; in other words, a change in Z is accompanied by a change in these parameters as well.

We purposefully model γ to be correlated with Z, yet as a distinct aggregate state, to isolate its impact on the economy. The agents in the economy understand that a recession will likely be accompanied by an increase in day traders, yet it is also possible to go through recessions without a change in the day traders. Furthermore, a recession in this economy will not just be a decline in average productivity but also a potential increase in the variance of productivity across firms and an increased level of fundamental uncertainty. Therefore, a change in γ isolates changes in the ability of stock prices to communicate the available information without conflating changes in the information available to the agents.

3 Measuring Price Informativeness

What are the real effects of liquidity shocks via information contained in stock markets? In this Section we measure *Price Informativeness (PI)*, and its cyclical behavior. We start by introducing the data sources and describing our empirical strategy. Then, we estimate the components to measure price informativeness (PI) in each year and each country, which we refer to as *a market*.[10] We leave the discussion of the model calibration and how to

[10] Hence, we suppress time subscripts. A market is then a collection of stocks that trade in stock markets of a particular country in a particular year. See Appendix B.1 for details.

recover the properties of liquidity shocks that are consistent with the model to the next Section.

3.1 Empirical Strategy and Data Details

3.1.1 Empirical Strategy

Price Informativeness (PI) can be obtained from equation (17). This can be done outside the full RBC model because traders are short-lived, and for all stocks that are traded in equilibrium, the pricing equation is independent of the behavior of households that are long-lived. The price informativeness series is granular and unrestricted in its cyclical behavior, as we let the pricing function parameters freely vary across markets (a country in a year), identifying them from cross-company variation. Using Assumption 1, the parameters of shock distributions ($\bar{\theta}_n, \bar{\theta}_d, \sigma^2_{\varepsilon_n}$, $\sigma^2_{\varepsilon_d}$, $\sigma^2_{\theta n}$, $\sigma^2_{\theta d}$) are firm invariant, and all heterogeneity across firms comes from shock realizations ($z_{in}, z_{id}, \theta_{in}, \theta_{id}, \varepsilon_{in}, \varepsilon_{id}$). Assumption 1 induces the equilibrium fractions of informed investors (λ_n, λ_d) and the pricing function parameters ($\phi_0, \phi_\varepsilon, \phi_n, \phi_d$) being firm invariant as well.

As is clear from equation (17), PI only depends on four parameters, the variances of the learnable part of profitability and liquidity of a stock ($\sigma^2_{\theta n}$ and $\sigma^2_{\theta d}$), and their price loadings (ϕ_n and ϕ_d). For this we need to construct the signals θ_n and θ_d to obtain the variances and to estimate ϕ_n and ϕ_d from the following stock pricing equation:

$$p_i = \phi_0 + \phi_\varepsilon \varepsilon^-_{in} + \phi_d \theta_{id} + \phi_n \theta_{in}. \tag{18}$$

where p_i can be measured for each company ticker i with the current stock price.

3.1.2 Data

We use data on firm-level stock prices of publicly traded firms from many countries, along with analyst forecast data (to measure expectations) and country-level economic conditions (to measure fundamentals).

Stock Prices and Fundamentals: We use Worldscope from Thomson/Refinitiv for data on monthly (open, close, high, low) stock prices and yearly fundamentals.[11] Consistent coverage starts around 1985 for the U.S. and by 2000 for most economies. Worldscope provides broad industry classifications to distinguish finance and insurance sectors, and historical exchange rates that allow historical and cross-company analyses.

Analyst Forecasts: We use the Institutional Brokers Estimate System (I/B/E/S) from Refinitiv for daily data on analyst forecasts of earnings-per-share.[12] We access both the Worldscope and the I/B/E/S through Wharton Research Data Services (WRDS), which provides a unique ticker for each company to link each company in the two datasets.[13]

Economic Conditions: For liquidity-related measures, we rely on banking crisis indicators by Baron et al. (2021) and continuous proxies of economy-wide liquidity from the World Bank. For measuring cyclicality, we rely on the weighted average of earnings of the publicly listed firms. See Table 3 for details.

3.2 Measuring Signals about Profitability and Liquidity

How do we measure the *signals* of profitability and liquidity observed by informed investors θ_{id} and θ_{in}? A common practice in the literature is to use realized values, i.e., z_{id} and z_{in}, as proxies for θ_{id} and θ_{in} (e.g., Bai et al. (2016), Dávila and Parlatore (2018)). However, according to our model without perfect information, regressing the price p_i on the realized values would lead to biased estimates of ϕ because the measurement error would be correlated with the realized earnings. We formalize this drawback in Proposition 3 for $\rho = 0$. The argument generalizes to $\rho > 0$, with more tedious algebra.

[11] Worldscope is a leading source of cross-country financials of publicly traded companies. It provides data from 98,000 companies in more than 120 countries, which amounts to 99% of the global market capitalization in 2021. Worldscope uses a standardized template for financial information that corrects for measurable differences in accounting practices across companies and markets. The entries are subject to automated tests that check accounting identities, outliers, and correlations for accuracy. See Reuters (2010) for details on standardization practices, accuracy tests, coverage, and selection criteria for companies.

[12] I/B/E/S collects forecasts about 22,000 active companies in 90 countries from over 18,000 analysts. Each observation is a forecast announcement made by an analyst regarding a balance sheet item of a company for a particular horizon. Forecasts are available for various payoff-relevant items, but those for Earnings-Per-Share (EPS) are the most widely available.

[13] We use the daily exchange rates provided by I/B/E/S to standardize the currency within a market across firms, time, and variables. See Appendix B.2 for details.

Proposition 3. *Let $\rho = 0$. An OLS regression of the price (p_i) on realized values (z_{id} and z_{in}) would give biased estimates of ϕ:*

1. $E[\hat{\phi}_0^B] = \phi_0 + \frac{\bar{\theta}_n \sigma^2_{\varepsilon_n} \phi_n}{\sigma^2_{\varepsilon_n} + \sigma^2_{\theta_n}} + \frac{\bar{\theta}_d \sigma^2_{\varepsilon_d} \phi_d}{\sigma^2_{\varepsilon_d} + \sigma^2_{\theta_d}}$
2. $E[\hat{\phi}_n^B] = \phi_n \left(1 - \frac{\sigma^2_{\varepsilon_n}}{\sigma^2_{\varepsilon_n} + \sigma^2_{\theta_n}}\right)$
3. $E[\hat{\phi}_d^B] = \phi_d \left(1 - \frac{\sigma^2_{\varepsilon_d}}{\sigma^2_{\varepsilon_d} + \sigma^2_{\theta_d}}\right)$

The bias becomes larger as the residual uncertainty after acquiring information ($\sigma^2_{\varepsilon_n}, \sigma^2_{\varepsilon_d}$) increases.

Given this drawback, we need to rely on signals. In other words, we should be interested in the extent to which prices reflect the traders' expectations at the time, not the eventual realizations.

The signal for future volatility, θ_{id}, is simply measured with the prior six months of realized price volatility, z_{id}.[14] For θ_{in}, this requires more work because we only observe forecasts, which, according to equation (2), is given by,

$$E[z_{in}] = Z + \theta_{in} + \rho\tilde{\varepsilon}_{in}$$

for an informed trader. We obtain $E[z_{in}]$ using one-step-ahead forecast for z_{in} from the median forecast in I/B/E/S for company i (announced within a 15-day window around the date the stock price is documented). How we decompose θ_{in} from $\rho\tilde{\varepsilon}_{in}$?

First, since we observe the time series $\varepsilon_{in} = z_{in} - E[z_{in}]$, we can compute its variance

[14] To estimate the stock price volatility, we use the measure proposed by Garman and Klass (1980), which only requires the opening (O), closing (C), highest (H), and lowest (L) prices during the time period. In particular, we look at the stock prices over the next six months to compute

$$\tilde{\sigma}^2_{it} = 0.511(H_{it} - L_{it})^2 - 0.019[(C_{it} - O_{it})(H_{it} + L_{it} - 2O_{it}) - 2(H_{it} - O_{it})(L_{it} - O_{it})] - 0.383(C_{it} - O_{it})^2$$

for each ticker i at date t. If the resulting volatility measure exceeds the stock price, we equate it to the stock price, imposing an intuitive limited liability rule on stock ownership. In Appendix C.2, we provide statistics on the cross-sectional distribution of the range volatility estimates and how they evolve over time for the median firm.

$\sigma^2_{\varepsilon_n}$ and obtain ρ from computing its autocovariance, since

$$\begin{aligned} Cov(F_t - E[F_t], F_{t+1} - E[F_{t+1}]) =& Cov(\theta_{int} + \rho\tilde{\varepsilon}_{in,t-1}, \theta_{in,t+1} + \rho(\rho\tilde{\varepsilon}_{in,t-1} + \tilde{\varepsilon}_{int})) \\ =& Cov(\rho\tilde{\varepsilon}_{in,t-1}, \rho^2\tilde{\varepsilon}_{in,t-1}) = \frac{\rho^3\sigma^2_{\varepsilon_n}}{1-\rho^2}. \end{aligned} \tag{19}$$

Second, once ρ is obtained, $\tilde{\varepsilon}_{in}$ can be approximated via perpetual inventory method using (2) and lagged values of ε_{in}.[15] Third, from the approximated series $\tilde{\varepsilon}_{in}$ we can finally obtain $\theta_{in} = E[z_{in}] - \rho\tilde{\varepsilon}_{in}$.

Table 1: Summary Statistics for the U.S. in 2015, Random Variables

variable	mean	sd	min	median	max
θ_n	0.06	0.06	-0.23	0.06	0.99
θ_d	0.21	0.14	0.05	0.17	1.00
ϵ_n	-0.01	0.04	-0.82	0.00	0.12
ϵ_d	0.02	0.10	-0.65	0.03	0.84
$\tilde{\epsilon}^-$	-0.01	0.05	-0.90	0.00	0.27

Notes: The figures are per unit of asset values constructed by multiplying original figures with the number of outstanding shares and dividing them by the value of their total assets.

As an illustration of results, Table 1 shows summary statistics for the series of signals and forecast errors that we estimated for the United States in 2015. In the Appendix, we provide further details for more years and for other countries.

3.3 Estimating Pricing Functions

Having measured the time series of signals about profitability and liquidity (θ_{id} and θ_{in}), we can proceed next to estimate the pricing loadings of these elements from the pricing equation (18).

[15] We first estimate ρ for markets that have at least 20 companies with past year's data. For the remaining markets, we use the country average if available and, if not, the overall average. For computing $\tilde{\varepsilon}_{in}$, we determine the lag order for each company-date pair based on the availability of past data on ε_{in}, with a maximum of three lags.

Adjustments to Accommodate our Assumptions: We make two adjustments before estimating the pricing functions. First, in the model, the stock-level shocks $(\theta_{in}, \theta_{id}, \varepsilon_{in}, \varepsilon_{id})$ are assumed to be independently distributed across firms. While providing tractability, this assumption rules out any correlation across stocks beyond the one driven by the aggregate shock. Second, to ensure Φ does not vary across stocks, in the model we assume the expected earnings ($\bar{\theta}_{in}$) and installed capital $\bar{K}_i$ do not vary across firms within a market. To accommodate these properties, we perform a factor analysis and residualize stock prices from common factors, past earnings, and total assets.

This factor analysis involves running the following regression for each stock i in market m at date t,

$$R_{it} = \alpha_i + \beta_{1i} MR_{mt} + \beta_{2i} SMB_{mt} + \beta_{3i} HML_{mt} + \varepsilon_{it} \tag{20}$$

using monthly observations from $t-23$ to t where $R_{it} = (p_{it} - p_{it-1})/p_{it-1}$. For each market-date pair, we construct the three Fama-French factors based on a balanced panel of stock prices from the past 24 months. The market return (MR) is constructed by looking at the month-to-month change in aggregate market cap in the market. The small-minus-big (SMB) is constructed as the difference in the aggregate returns of the top and bottom 30% stocks in terms of market cap. The high-minus-low (HML) is constructed as the difference in the aggregate returns of the top and bottom 30% stocks in terms of book-to-market ratio. We then use the estimates for β_{1i}, β_{2i}, and β_{3i}, which represent the factor loadings (the 'betas') for firm i, to residualize the prices by regressing them on second-order polynomials of the estimated betas, latest eps announcement (representing $\bar{\theta}_{in}$), and total assets (representing $\bar{K}_i$),.

Second, in the model, we assume each stock provides ownership of one unit of installed capital in the firm. In the data, however, the meaning of a single share, hence the stock price and earnings-per-share (eps), differ across firms. To make these variables comparable across firms and consistent with our model, we transform the factor-adjusted stock prices, eps, and eps forecasts to per-unit-of-asset values. We do this by multiplying the original value by the number of outstanding shares of the firm during a year and dividing it by the value of the Total Assets reported at the end of that year.

Timing: The data sources we use are of varying frequencies, and the accounting years (AY) differ across firms. This introduces several timing challenges. First, while data on stock prices is monthly, data on company fundamentals is yearly. Second, flow variables, such as earnings, refer to flows during the AY, while stock variables, such as total assets, refer to values at the end of the AY.[16] Third, the eps forecast announcements are available daily even though the relevant target dates, by construction, are yearly.

To tackle these challenges, we consider the stock price for each stock i six months before the end of the respective firm's AY. Call this date D_{it} where t refers to the associated year. For each stock-year pair it, we use the stock price at D_{it} to represent the model object p_i. Next, we map the median of the analysts' forecasts that are announced within a 15-day band around D_{it} for the current year eps to $Z + \theta_{in} + \rho\tilde{\varepsilon}_{in}^{-}$. The realized value for the same eps is then mapped to $Z + \theta_{in} + \rho\tilde{\varepsilon}_{in}^{-} + \varepsilon_{it}$.[17]

Figure 1 provides an example for a firm i whose previous AY ended in April 1995. Then, p_i is measured on October 1995. For θ_{in}, we use the forecasts that are announced around October 1995 for the earnings during the AY that ends in April 1996 (eps_i^f). The latest announcement for earnings (eps_i^a) on April 1995 represents what's publicly known when p_i is determined. For θ_{id}, we look at the realized range volatility between April 1995 and October 1995 ($Range_i$). The realized range volatility between October 1995 and April 1996 provides z_{id}.

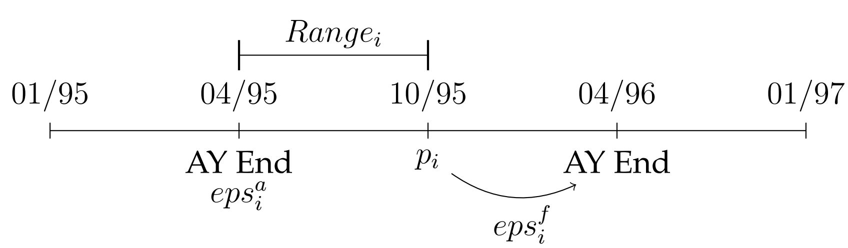


Figure 1: A Timing Example

This choice has implications for the interpretation of coefficients. Since the end of the accounting year is Dec 31st for most companies, the yearly estimates will generally refer to the stock prices and forecasts around July of the corresponding year. Hence, the effects

[16]Furthermore, the accounting year generally differs across firms, and fundamentals are publicly announced a couple of months after the last day of the accounting year.

[17]If the stock prices were sampled at the same date for all firms, the traders' information set would differ firm by firm. Instead, we sample prices at different points in time to make sure that *i)* prices are equally spaced within the respective firm's AY and *ii)* the previous year's fundamentals are already announced, i.e., the stock prices reflect traders' knowledge of $\tilde{\varepsilon}_{in}^{-}$. If no earnings forecasts are available at six months prior, we use five and seven months prior, in that order.

of a major event before July, e.g., Bernanke's 'Taper Tantrum' in May 2013, are expected to be seen in the estimates for 2013. On the other hand, the effects of a major event after July, e.g., the Black Monday sell-off in October 1987, is expected to be seen in the estimates for 1988.

Sample Restrictions: First, to guarantee an unambiguous match between the relevant monthly stock prices and the yearly fundamentals for each firm, we remove observations for which *i)* firms are ever cross-listed in multiple stock exchanges or in the finance/insurance sectors, *ii)* the listed accounting year-end dates are inconsistent (more than 12 months ahead) with the date of the stock price, and *iii)* the financial statements are announced earlier than the end of the reported accounting year or after more than six months. Second, to avoid firms that promise short-run losses with a possibility of abnormally high earnings in the long run, we remove observations where the company's earnings forecast indicates losses larger than 10% of the total value of its assets.[18] Third, to run the Fama-French analysis, we need monthly stock prices available 6 months before and after and we require the associated market to have more than 30 stocks that constitute a balanced panel with at least 12 months of stock price data available in the past 24 months. Finally, we winsorize the adjusted earnings forecasts, earnings, and stock prices at a 5% level to deal with stock anomalies and potential inaccuracies in data.

Estimating Pricing Coefficients Based on the data adjustments described above, here we estimate the pricing function (18) for each market. We are particularly interested in the estimates for ϕ_n and ϕ_d that enter into the calculation of Price Informativeness ($R(\theta_{in}, E[\theta_{in}|p_i])$) as derived in (17).

We estimate the pricing function (18) in two steps. We first residualize stock prices as described above to i) capture the heterogeneity in $\bar{K}_i$ and $\bar{\theta}_{in}$ that is not captured in the model and ii) capture dependencies in the distributions of θ and ε across companies. Then, we estimate the following pricing function using the residualized stock prices for each market:

$$\hat{p}_i = \beta_0 + \beta_1 \tilde{\varepsilon}_i^- + \beta_2 Range_i + \beta_3 e\hat{p}s_i^f + \upsilon_i. \tag{21}$$

where $\tilde{\varepsilon}_i^-$ is estimated using forecast errors obtained in Section 3.1.1, $Range_i$ is the

[18] These observations are predominantly pharmaceutical companies that run consistent losses for several years. The correlation between earnings and stock price forecasts is negative for these firms, while it is close to 1 for the rest of the sample.

range volatility estimate for the stock price in the past six months and is used as the empirical counterpart of θ_{id}. $\hat{eps}_i^f$ is the forecast for the next announcement of the earnings adjusted for known components and is used as the empirical counterpart of θ_{in}. Under Assumption 1, the OLS estimator for β_2 and β_3 provide unbiased estimates of ϕ_d and ϕ_n.[19]

As an illustration of the results, Table 2 shows summary statistics for the series we used in this section to estimate pricing functions for the U.S. in 2015.

Table 2: Summary Statistics for the U.S. in 2015, Pricing Function

variable	mean	sd	min	median	max
ALPHA	0.00	0.03	-0.09	0.00	0.77
BMarketReturn	1.08	1.00	-18.53	1.00	6.79
BSMB	0.10	0.37	-1.75	0.08	5.69
BHML	-0.04	1.52	-45.22	-0.16	7.13
p	1.52	1.25	0.30	1.12	6.14
$\bar{K}_i$	11,709.18	34,237.83	19.85	2,122.20	552,257.00
$\bar{\theta}_n$	0.05	0.06	-0.57	0.05	0.15

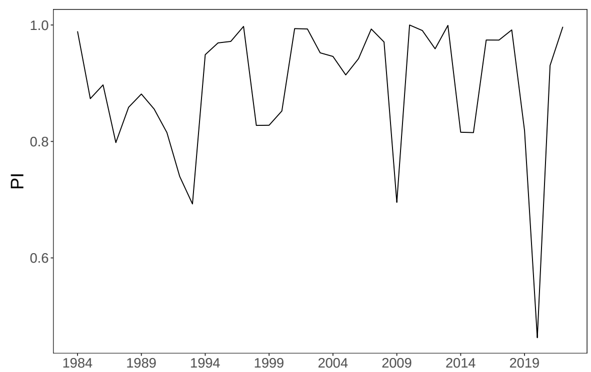
Notes: The p and $\bar{\theta}_n$ figures are per unit of asset values constructed by multiplying original figures with the number of outstanding shares and dividing them by the value of their total assets.

3.4 Price Informativeness Over Time and Across Countries.

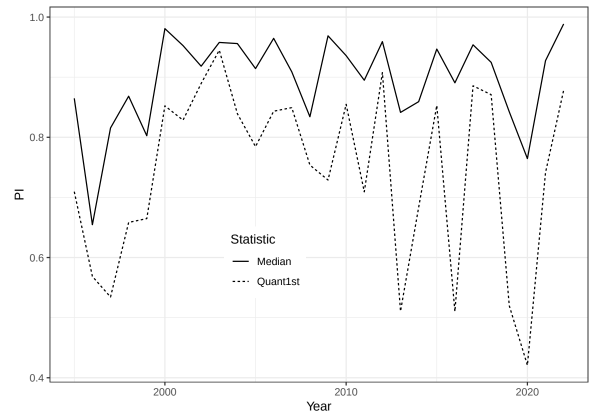
We have determined the four parameters, $\sigma^2_{\theta n}$, $\sigma^2_{\theta d}$, ϕ_n and ϕ_d that we need to compute price informativeness (PI) from equation (17). We do this in 381 country-year observations from 22 countries. The U.S. has data starting in 1984, and many other countries starting in the 1990s.[20] As an illustration, Figure 2a presents the price informativeness series for the U.S., which faced its largest price informativeness decline in our 40-year sample during the COVID-19 episode, the 2008-09 financial crisis (the so-called Great Recession), and the 1990-1991 recession. Right after these episodes, price informativeness went back to its

[19] The estimates are in line with the model's predictions. The model suggests $\phi_n > 0$ and $\phi_d < 0$. In all but one market, $\hat{\phi}_n > 0$ and in 70% of the markets $\hat{\phi}_d < 0$ (100% and 90% once we restrict attention to estimates significantly different from zero at 5% level.). Furthermore, the model predicts $\frac{|\phi_n|}{|\phi_d|} > 1$ which is true for 97% of the markets.

[20] For countries in the Eurozone, we restrict attention to 1999 onwards due to a lack of reliable currency conversion between their old and new currencies.



(a) Price Informativeness Estimates for the U.S.



(b) Cross-Sectional Statistics for PI Estimate

Figure 2: The PI Estimates

Notes: We restrict attention to countries for which a PI is estimated for at least 20 years between 1994 and 2022 to get a partially balanced sample. The countries are Australia, Germany, United Kingdom, France, Japan, Taiwan, Canada, Sweden, and the U.S. See Figure 14 in Appendix D for a box plot of global estimates from 1984 and 2021.

pre-crisis level.

The drivers behind each decline, however, were different. The four components that go into the computation of the PI measure are presented in Figure 3. On the one hand, variances changed in these periods. While both $\sigma^2_{\theta n}$ and $\sigma^2_{\theta d}$ increased around the Great Recession and the COVID-19 pandemics, only $\sigma^2_{\theta d}$ increased during the 1990-1991 recession. On the other hand, the pricing sensitivity to earnings, ϕ_n, only saw a noticeable decline during the Covid pandemic, while the sensitivity of prices to liquidity, ϕ_d, increased in all crises. To provide a more systematic decomposition of these drivers, we ask: How much would the time series variance of the PI measure decline if the component x is kept fixed at its median value? The variance would decline by 4%, 21%, and 71% with $\sigma^2_{\theta d}$, ϕ_n, and ϕ_d at their median vales, respectively. If $\sigma^2_{\theta n}$ was kept at its median value, the variance would increase by 22%.[21] In other words, the large declines in PI largely follow rises in the correlation between prices and volatilities and, partly, dips in the correlation between prices and earnings forecasts.

Figure 2b presents moments from the yearly distribution of Price Informativeness estimates across 17 countries between 1995 and 2021. The median and the 1st quantile

[21] This is, of course, just a statistical decomposition, not a structural one, as ϕ_n and ϕ_d are themselves functions of $\sigma^2_{\theta n}$ and $\sigma^2_{\theta d}$.

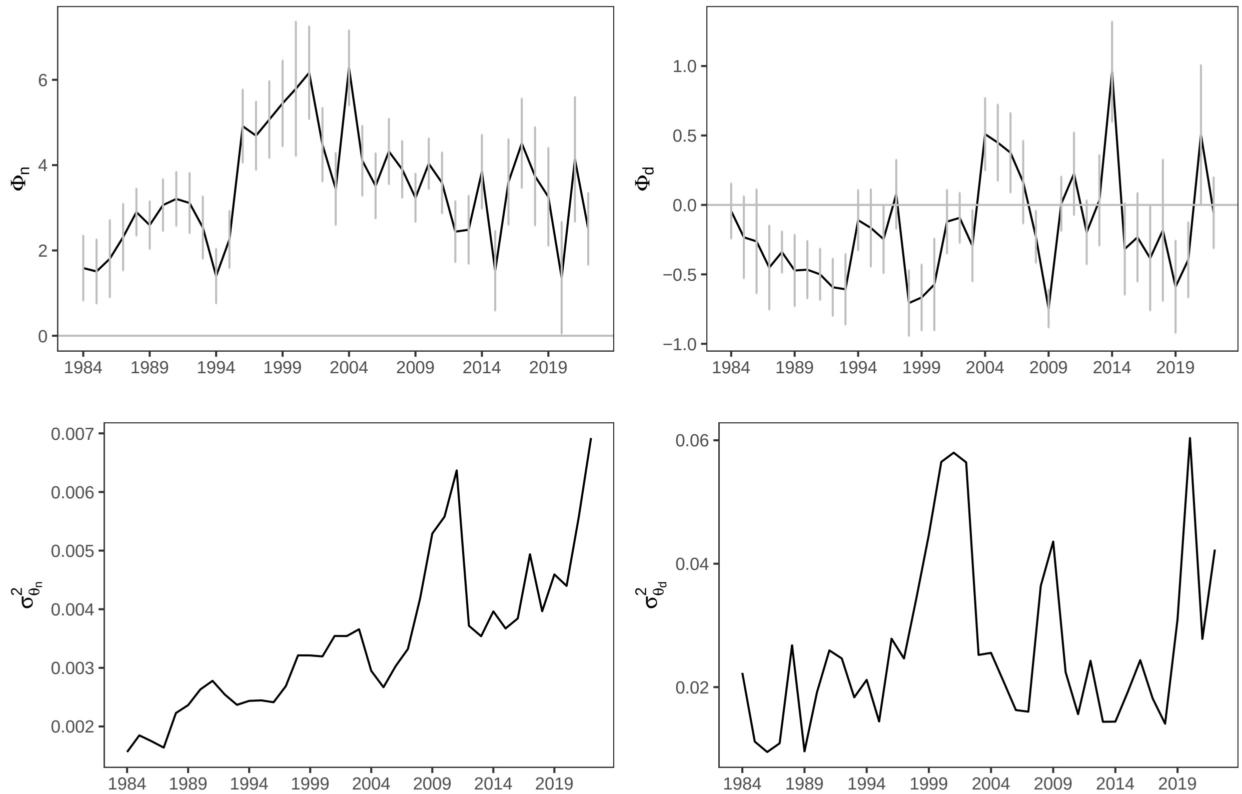


Figure 3: Estimated Price Informativeness Components for the U.S.

experienced declines around the Great Recession and the COVID-19 pandemic, similar to the U.S. On the other hand, there was no major international reaction to the burst of the dot-com bubble. In contrast, major declines were observed in Price Informativeness in *i)* Australia and Korea around the Asian Financial Crisis in 1997, *ii)* most European economies around the 'Taper Tantrum' and the European sovereign debt crisis in 2013, and *iii)* UK and France around the Brexit decision in 2016.

The pattern that PI declines during recessions that we highlighted in Figures 2a and 2b is, however, only suggestive. Table 4 tests this hypothesis formally. We estimate:

$$PI_{it} = \beta_0 + \beta_1 Z_{it} + \beta_2 y_{it} + F_{ci} + F_{yi} + \varepsilon_{it}, \tag{22}$$

where Z_{it} represents the aggregate productivity process of publicly listed firms and is measured by the weighted average of normalized earnings. PI_{it} is price informativeness estimated from (21). y_{it} represents several proxies of the extent to which stock markets are used to face liquidity needs in country i and year t. For this variable, we use binary crisis measures and continuous proxies. For binary measures, we use the presence of banking panics and banking equity crises as constructed by Baron et al. (2021). For continuous measures, we use capital-asset ratios of the banking sector, loan spreads of the banks (lending rate minus treasury bill rate), and the ratio of nonperforming loans to total gross loans. The capital asset ratio is expected to be low, while the loan spreads and non-performing ratios are expected to be high in low-liquidity environments, such as the Great Recession. Finally, F_{ci} and F_{yi} are country and year fixed effects.

Table 3: Summary Statistics on Economic Conditions in A Panel of Countries

Statistic	N	Mean	St. Dev.	Min	Max
Bank Capital to Assets	188	6.671	1.965	4.109	10.565
Bank Loan Spreads	197	3.944	6.450	−0.032	39.216
Non-performing Loans	193	2.454	2.449	−0.090	16.911
Banking Panic	265	0.038	0.191	0	1
Banking Equity Crisis	265	0.034	0.181	0	1
PI	371	0.812	0.241	0.001	1.000
Avg Earnings	371	0.054	0.016	0.017	0.097

Notes: The average earnings denote the weighted average of the normalized earnings. The liquidity crisis indicators, Banking Panic and Banking Equity Crisis are from Baron et al. (2021). The continuous liquidity measures, Bank Capital to Asset, Bank Loan Spreads, and Non-performing Loans are from the World Bank. The authors estimate the Price Informativeness (PI) measure. The even-numbered columns have country-fixed effects. The standard errors are clustered at the country level. *p<0.1; **p<0.05; ***p<0.01

Table 4 presents the results. Across different specifications lower economy-wide liquidity is associated with lower levels of price informativeness, consistent with our theory. A one standard deviation increase in the capital-to-asset ratio of the banking system, for example, is associated with a 0.2 increase (0.83 standard deviations) in Price Informativeness. Importantly, higher aggregate productivity, once conditioned on the level of economy-wide liquidity, has a much smaller impact on price informativeness, which is not statistically significant.

In the next Section, we use the estimated fluctuations in Price Informativeness to infer the structural parameters that constitute the sources of those fluctuations.

Table 4: Price Informativeness and Economic Conditions

	PI				
	(1)	(2)	(3)	(4)	(5)
Avg Earnings	−0.16	−2.17	0.22	−2.59	−2.58
	(0.10)	(2.59)	(2.64)	(2.07)	(2.08)
Bank Capital to Assets	0.10***				
	(−0.02)				
Bank Loan Spreads		−0.02*			
		(0.01)			
Non-performing Loans			−0.03**		
			(0.01)		
Banking Panic				−0.16*	
				(0.09)	
Banking Equity Crisis					−0.05
					(0.09)
Range	2005-2022	1984-2022	2005-2022	1984-2016	1984-2016
Country & Year FE	Yes	Yes	Yes	Yes	Yes
Observations	188	197	193	265	265

Notes: The standard errors are clustered at the country level. See Table 3 for the details on the variables. *p<0.1; **p<0.05; ***p<0.01

4 Model Calibration

In this section, we calibrate the model parameters to the publicly traded companies of the U.S. economy. Later, we use the calibrated model to quantify the contribution of changes in the composition of traders (γ) on misallocation through its impact on the price informativeness of stock markets and investigate the role of endogenous information acquisition.

In the previous section, we have estimated the time series of the means ($\bar{\theta}_n, \bar{\theta}_d$) and the variances ($\sigma^2_{\theta_n}$, $\sigma^2_{\theta_d}$, $\sigma^2_{\varepsilon_n}$, and $\sigma^2_{\varepsilon_d}$) associated with the stock returns. In this section, we finalize the model's parameterization in two steps. First, we externally calibrate standard parameters. Second, we simultaneously discretize the aggregate state, estimate the information cost functions, and estimate the series of composition of traders.

4.1 Calibration Strategy

Externally Calibrated Parameters For households, we set the intertemporal elasticity of substitution $1/\eta$ equal to 0.5. For traders, following Farboodi and Veldkamp (2020), we set the absolute risk aversion parameter a equal to 0.05. Additionally, we set the depreciation rate δ equal to 0.1, the risk-free interest rate available to the traders r^F equal to 0.02, and the autocorrelation of the unlearnable part of profitability, ρ, to 0.[22]

Discretizing the Aggregate States We have to estimate the process of the aggregate state defined as $s = \{Z, \gamma\}$. We first remove a linear trend from the original series of the logarithm of earnings per share Z and liquidity needs γ. Then, we estimate a vector-auto-regression and discretize the estimated vector-auto-regression into a 4-state Markov chain (2 levels for each state) following Gospodinov and Lkhagvasuren (2014). We assign each year as a high or low Z given the estimated Z grid and compute the average levels of $\sigma^2_{\theta_n}$, $\sigma^2_{\theta_d}$, $\sigma^2_{\varepsilon_n}$, and $\sigma^2_{\varepsilon_d}$ in those years. These averages constitute their levels in high and low Z states in the model.

Series of Liquidity Needs Equation 16 provides a relationship between the ratio of the pricing coefficients and liquidity needs, captured by the fraction of daily traders, γ. This relation depends on $Var(z_{in} + p'_{in})$ and λ_n/λ_d.[23]

Estimating $Var(z_{in} + p'_{in})$ is nontrivial because part of the uncertainty is about future prices and hence future pricing coefficients. Since the pricing function shows persistence, the conditional expectation also differs from the unconditional one. To estimate the conditional variance of the future payoff, we estimate a first-order auto-regressive process for the estimated pricing coefficient vector $\hat{\Phi}$:

$$\hat{\Phi}_t = B\hat{\Phi}_{t-1} + U + W_t, \tag{23}$$

where $\hat{\Phi}_t = [\hat{\phi}_0 \; \hat{\phi}_\varepsilon \; \hat{\phi}_n \; \hat{\phi}_d]$, B is a 4×4 diagonal matrix that controls the persistence,

[22] If $\rho \neq 0$ the $\tilde{\varepsilon}_i$ never reaches an ergodic distribution under aggregate shocks, and its distribution becomes an additional state variable. Even though accounting for a potential correlation is necessary for interpreting the data, it wouldn't play a significant role in our model due to the static allocation of capital across companies.

[23] The only component of $z_{in} + p'_{in}$ that is persistent is $\tilde{\varepsilon}^-_{in}$ and its stochastic component ε_{in} is *iid* across stocks and over time. Hence, $Var(z_{in} + p'_{in})$ does not vary across i.

U is a 4×1 vector that stores the constant terms, and W_t is a 4×1 error term where $W_t \sim MVN(0, Q)$ and Q is the associated variance-covariance matrix. Once estimated, (23) provides a simple formula for $Var(z_{in} + p'_{in})$.

After recovering $Var(z_{in} + p'_{in})$, we use Corollary 2 to back out λ_n/λ_d. In particular, the following equations hold in an interior solution to the information acquisition problem.

$$e^{ac_n(\lambda_n)} \operatorname{Var}\left(\epsilon_{in} + p'_i\right) = \operatorname{Var}\left(\epsilon_{in} + p'_i\right) + \frac{\sigma^2_{\theta_n}}{1 + \frac{\sigma^2_{\theta_n}}{\sigma^2_{\theta_d}}\left(\frac{\phi_n}{\phi_d}\right)^2} \tag{24}$$

$$e^{ac_d(\lambda_d)} \left(\sigma^2_{\varepsilon_d} + \operatorname{Var}\left(\epsilon_{in} + p'_i\right)\right) = \left(\sigma^2_{\varepsilon_d} + \operatorname{Var}\left(\epsilon_{in} + p'_i\right)\right) + \frac{\sigma^2_{\theta_n}}{1 + \frac{\sigma^2_{\theta_n}}{\sigma^2_{\theta_d}}\left(\frac{\phi_n}{\phi_d}\right)^2} + \frac{\sigma^2_{\theta_d}}{1 + \frac{\sigma^2_{\theta_d}}{\sigma^2_{\theta_n}}\left(\frac{\phi_d}{\phi_n}\right)^2} \tag{25}$$

An advantage of our setting is that equations 24 and 25 allow recovering λ_n and λ_d *for a given* cost function $c(.)$. Using the estimated values for λ_n/λ_d and $Var(z_{in} + p'_{in})$, and (16), we can recover the time series for γ. The challenge, however, is to discipline the information cost functions $c(.)$. We next discuss our strategy.

Information Cost Functions We parametrize the information cost function as follows:

$$c_j(\lambda_j) = \nu_j \left(\frac{1}{1 - \lambda_j}\right)^{\psi_j} - \nu_j \tag{26}$$

for $j \in \{d, n\}$, which satisfies $c_j(0) = 0$ and $\lim_{x \to 1} c_j(x) = \infty$. This functional form simplifies the equilibrium computation by mapping each nonnegative cost value to a unique $\lambda_j \in [0, 1)$. For each set of parameters, $\{\nu_d, \nu_n, \psi_d, \psi_n\}$, the stock trading module provides a mapping between the pricing parameters (ϕ) and the fraction of informed agents (λ) through equations (13), (24) and (25). Our estimation strategy relies on choosing these four parameters to match the moments of the pricing function in booms and busts. In particular, we focus on seven moments implied by the estimated pricing function: (1) the level of price informativeness in low productivity periods, (2) the change in price informativeness after aggregate shocks (3) the average fractions of day and night traders that are informed, and (4) the growth in the information acquisition activities during busts.[24]

The level of the cost of acquiring information depends on ν_d and ν_n, while its curvature is determined by ψ_d and ψ_n. The average levels of λ_n and λ_d are hence directly informative about ν_d and ν_n, together with the average level of price informativeness, which depends on λ_n and λ_d. The changes in the level of price informativeness and information acquisition activities, on the other hand, tell us the extent to which day and night traders respond to economic changes: a low ψ_d (ψ_n) makes it easier for a day (night) trader to scale their information acquisition, tilting the prices to reflect more of θ_d (θ_n).

Estimation Algorithm We start with a guess $\{\nu_d^0, \nu_n^0, \psi_d^0, \psi_n^0\}$ and take the next steps:

1. Start iteration k with $\{\nu_d^k, \nu_n^k, \psi_d^k, \psi_n^k\}$. Use (24) and (25) with estimated series for pricing coefficients and parameters to infer $\lambda_{d,data}^k$ and $\lambda_{n,data}^k$ series.
2. Invert (16) to infer γ series.
3. Follow the discretization procedure discussed above to estimate a vector-auto-regression for $\{Z, \gamma\}$ and discretize it into a Markov Chain. Compute the values for $\sigma_{\theta_n}^2$, $\sigma_{\theta_d}^2$, $\sigma_{\varepsilon_n}^2$, and $\sigma_{\varepsilon_d}^2$ associated with each Z level.
4. Compute the stock market equilibrium and simulate the economy for the model moments M_{sim}^k (See Table 6 for a list of moments).
5. Compare M_{sim}^k with M_{data}^k. If the discrepancy is below the threshold, stop. If not, go back to step 1 with new $\{\nu_d^{k+1}, \nu_n^{k+1}, \psi_d^{k+1}, \psi_n^{k+1}\}$.

After having estimated the cost function parameters, we estimate the discount factor to generate a 2% risk-free interest rate.[25] Because we directly estimated the level of productivity from earnings data, we have a degree of freedom to set the scale of the economy. So, we normalize the adjustment cost $\bar{k}/\xi$ to achieve an average capital level of 1.

[24] To be precise λ_d and λ_n are not moments observed directly from the data. However, they can be inferred from the estimated pricing function for a given cost function. Given the estimated parameters, matching the price informativeness boils down to matching ϕ_n/ϕ_d. The reason why the ratio doesn't exactly pin down λ_d and λ_n is the $Var(z_{in} + p_i')$ term, which depends on the entire pricing function and how it changes from booms to busts. In other words, the estimated parameters of the pricing function in booms and busts provide the two missing moments in this estimation.

[25] This interest rate is not the return households receive from capital (r) since the realized return accrues all the surplus to the households. Instead, we evaluate the counterfactual interest rate in the model that would clear the markets in a competitive setting.

Table 5: Estimated Aggregate State Levels

s	Z	γ	σ_{θ_n}	σ_{θ_d}	σ_{ϵ_n}	σ_{ϵ_d}	Years
1	0.051	0.38	0.058	0.18	0.036	0.14	'96,'97, '01, '02, '13, '16
2	0.051	0.8	0.058	0.18	0.036	0.14	'89, '90, '91-'94, '98-'00, '03, '09, '12, '15, '17, '19-'21
3	0.063	0.38	0.059	0.14	0.038	0.12	'84,'95,'04,'07,'08,'10,'11,'18,'22
4	0.063	0.8	0.059	0.14	0.038	0.12	'85-'88,'05,'06,'14

Notes: The last column shows the years with Z and γ estimates closest to the values associated with each aggregate state. The values of $\sigma^2_{\theta_n}$, $\sigma^2_{\theta_d}$, $\sigma^2_{\varepsilon_n}$, and $\sigma^2_{\varepsilon_d}$ are estimated by taking the averages over the years associated with the Z values.

4.2 Calibration Results

Table 5 summarizes the estimates of the four aggregate states and the standard deviations of signals and forecast errors in each state that result from our discretization exercise. The Z value fluctuates between 0.0507 and 0.0627 while γ fluctuates between 0.38 and 0.8. The values for Z are consistent with median earnings per share fluctuations over time. The estimated γ process indicates that roughly two-thirds of the traders focus exclusively on the long-run earnings of stocks during high liquidity periods, while about 80% consider the short-run price fluctuations when liquidity dries up. The table also assigns the years that the estimation assigns to each aggregate state.

Table 6 summarizes all calibrated parameters, both externally and internally. We also summarize the changes in price informativeness across the four aggregate states. We set as a benchmark price informativeness ($\underline{PI}$) the one corresponding to the aggregate state $(\overline{Z}, \underline{\gamma})$, which we interpret as "normal times". Consistent with data, everything else constant, price informativeness declines when the economy transitions to a state with higher liquidity needs and increases when the economy transitions to a state with lower productivity. If both productivity declines and liquidity needs increase, the second effect is more prevalent and price informativeness declines.

Figure 4 demonstrates the estimated shape of the estimated cost functions for day and night traders, together with the median levels of λ_d and λ_n (shown in Table 6). The estimated cost function indicates that the cost of learning about the liquidity of stocks is much higher and steeper than the cost of learning about their fundamental payoffs. This result is disciplined by two observations: (*i*) the post-information acquisition uncertainty in stock liquidity ($\sigma^2_{\epsilon_d}$) is much higher than the uncertainty in fundamental payoffs

Table 6: Externally and Internally Calibrated Parameters

Parameter	Value	Moment	Parameter	Value	Moment	Model	Target
$\bar{k}/\xi$	600	Normalized	ν_n	4.22	λ_d	0.06	0.02
η	2	External	ν_d	0.12	λ_n	0.29	0.33
a	0.05	External	ψ_n	0.27	$\underline{PI}$	0.87	0.87
δ	0.1	External	ψ_d	58.6	$\Delta PI_{\overline{z}\underline{\gamma}\to\underline{z}\underline{\gamma}}$	0.02	0.01
					$\Delta PI_{\overline{z}\underline{\gamma}\to\overline{z\gamma}}$	-0.16	-0.11
					$\Delta PI_{\overline{z}\underline{\gamma}\to\underline{z}\overline{\gamma}}$	-0.10	-0.16
					$\Delta\lambda$	0.17	0.22
r^F	0.02	External	β	0.955	Real Interest Rate	0.02	0.02

Notes: λ_d and λ_n are computed as the average values in the ergodic distribution of the aggregate states. Their data counterpart is computed as the median value in the inferred series. $\underline{PI}$ is computed as the price informativeness level averaged over low Z states in the ergodic distribution. Its data counterpart is computed as the average PI level in designated low Z years. The ΔPI terms are computed as percentage changes in PI measures in an IRF exercise where the economy moves from a long sequence of high Z low γ states. Their data counterpart is computed as the percentage differences in average PI levels across years with different aggregate state designations. The $\Delta\lambda$ is computed as the percentage change in aggregate (day and night) fraction of informed traders in an IRF exercise where the economy moves from a long sequence of high Z low γ states to a state of low Z high γ. Its data counterpart is the percentage difference in employment in the NAICS 52394 (Portfolio management and investment advice) industry during recessions and booms (CES Survey, BLS).

($\sigma^2_{\epsilon_n}$), and (*ii*) the price uncertainty c($Var(\epsilon_{in} + p'_i)$) is much larger than the dispersion of signals ($\sigma^2_{\theta_n}$ and $\sigma^2_{\theta_d}$). As a result, (24) and (25) indicate that the night traders must be paying a higher cost. Under symmetric cost functions, paying a higher cost would imply more information acquisition by night traders and a high price informativeness. However, this scenario would be inconsistent with the information acquisition motives of the day traders in equilibrium: day traders should be willing to pay more than night traders, given the additional uncertainty they face. Furthermore, the increase in information acquisition in recessions would be as large as 200% as new day traders would acquire much more information than the old night traders. Hence, the cost function should be more restrictive for day traders to match the level of price informativeness and be steep enough to prevent large swings in information acquisitions between booms and busts.

Figure 5 provides the γ series implied by the estimated cost functions depicted in Figure 4. The first panel shows the estimated liquidity needs series for the U.S. These suggest elevated transitory liquidity needs, γ, around the Great Recession and the COVID-19 pandemics, and a persistent period of high liquidity needs from 1985 to 1995, with a spike right after the Black Monday stock market crash. While the average fraction of traders

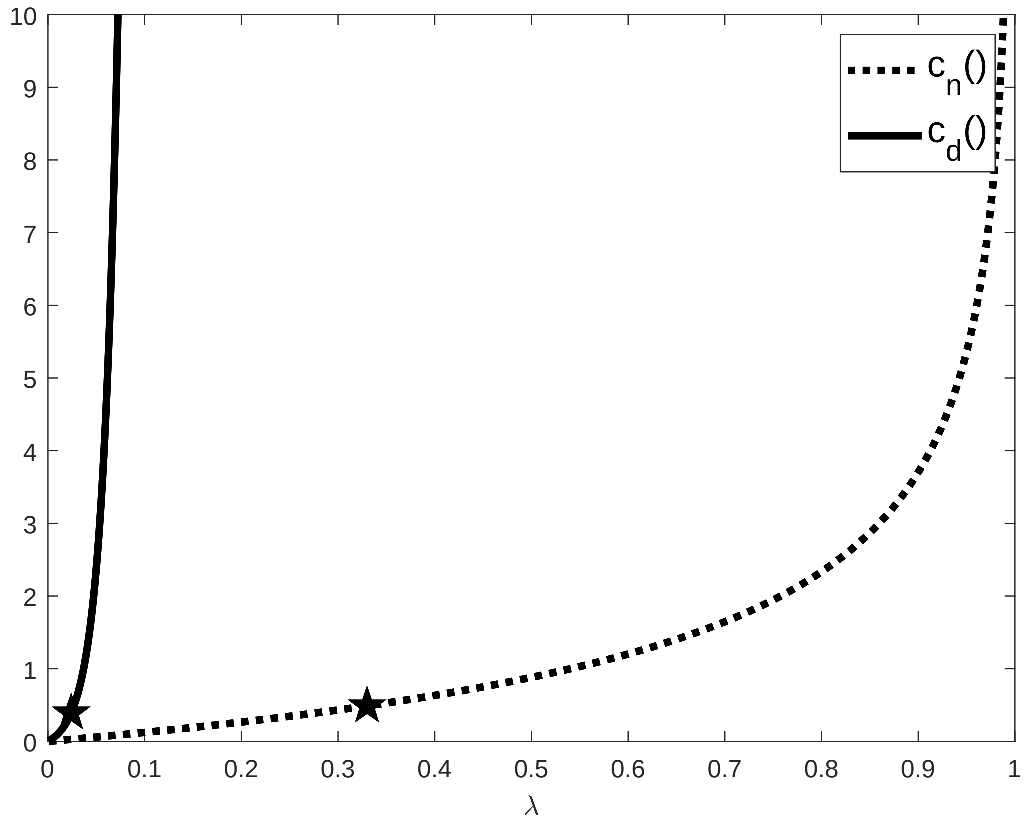


Figure 4: Estimated Information Acquisition Cost Functions
Notes: The solid and the dashed lines represent $c_n()$ and $c_d()$, respectively. The stars refer to the median levels of λ and associated costs in the stochastic steady state.

with liquidity needs after 2000 ranges around 40%, these three events reach levels above 80%. This is consistent when considering other developed countries. In the second panel of the figure we show the estimated series of liquidity needs for the median of developed economies using the estimated information costs for the U.S. Naturally, these series are less volatile than for a single country but display similar patterns of increase during stressed periods, like the COVID-19 pandemics or the European Sovereign Debt crisis.

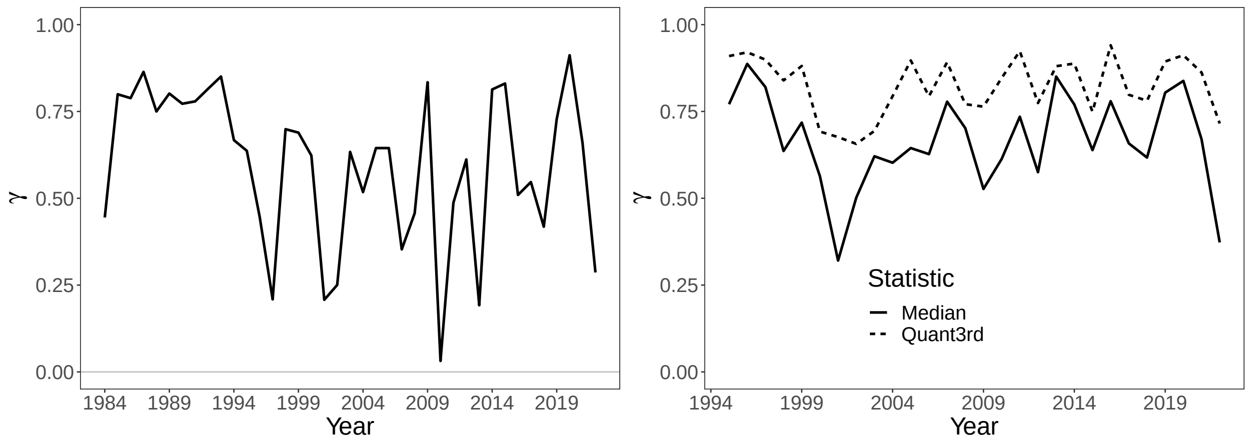


Figure 5: γ Series Implied by the Estimated Cost Function Notes: The left panel presents the estimated γ series for the United States, while the right panel presents the cross-sectional moments from a panel of countries using the cost function estimate from the U.S. To get a partially balanced sample, we restrict attention to countries for which a PI is estimated for at least 20 years between 1994 and 2022. The countries are Australia, Germany, United Kingdom, France, Japan, Taiwan, Canada, Sweden, and the U.S.

5 Quantitative Relevance of Stock Price Informativeness

In this section we use the calibrated model to assess the quantitative relevance of stock price informativeness on aggregates via its effect on resource misallocation. The first exercise is to estimate an impulse response function for a recessionary shock with and without a liquidity shock. This comparison means to capture recessions with and without distress in financial markets, and informs us how an increase in traders' liquidity needs amplify or dampen the impact of a recessionary shock through changes on stock price informativeness. The second exercise compares the aggregate effects of recessions with financial distress with alternative economies, one with different information costs and another in which traders receive information exogenously. While the first economy is useful to evaluate disclosure policies, the second shows the quantitative importance of modeling information acquisition as a choice.

5.1 Allocation Effects of Productivity and Liquidity Shocks

To assess the quantitative relevance of liquidity needs on aggregates through their impact on stock price informativeness and the corresponding misallocation of resources, we simulate our economy for a long time with a high Z and low γ (state 3 in Table 5).[26] Then, we introduce a one-period recession *with* financial distress: a decline in Z *and* an increase in γ (a transition to state 2 in Table 5). This *dual shock* captures the major downturns of the U.S. economy during our estimation period: the Great Recession in 2009 and the COVID-19 recession in 2020. We then compare these results to the aggregate effects of a recession *without* liquidity distress: a drop in Z that is *not* accompanied by an increase in γ (a transition to state 1 in Table 5). This *single shock* captures recessions without clear liquidity problems, such as in 2001. Figure 6 presents the results of these two different aggregate shocks: the solid line in the first case and the dotted line in the second case.

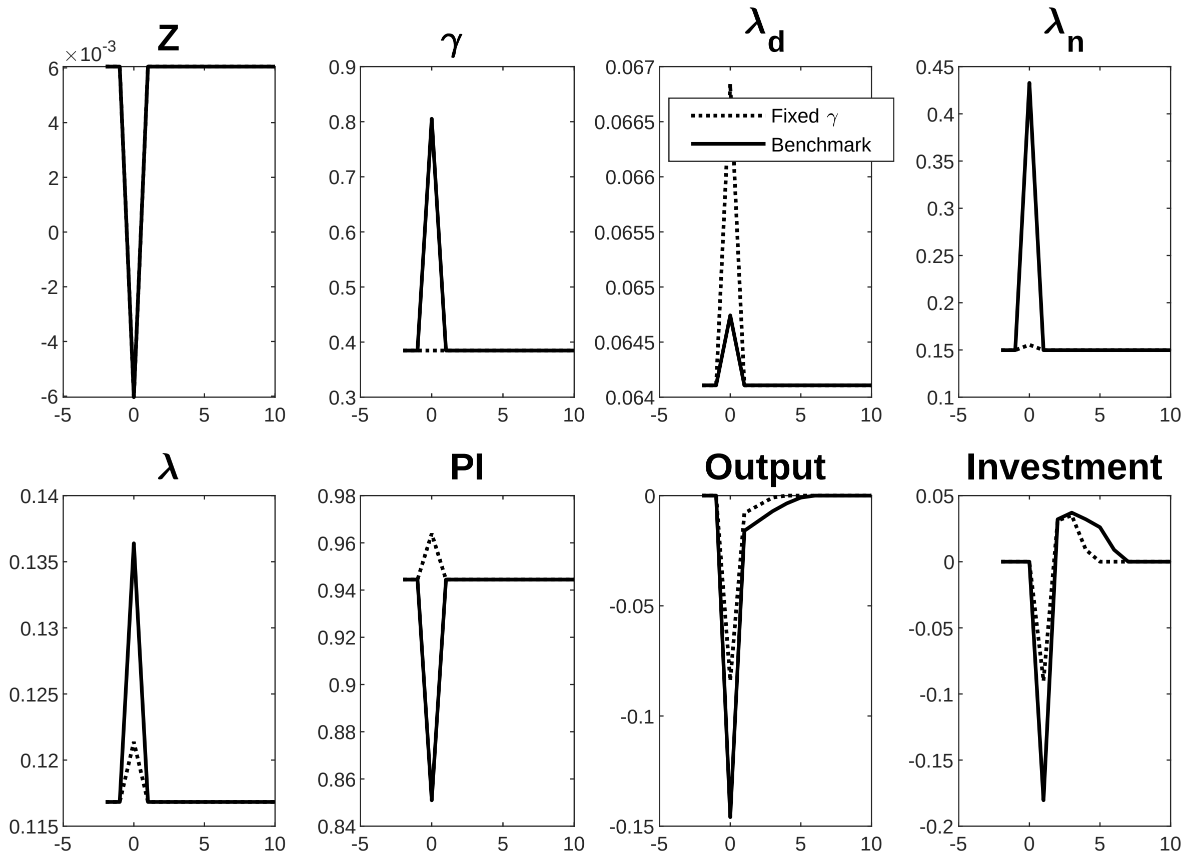


Figure 6: Impulse Response Functions
Notes: The first two panels provide the shocks that hit the economy under the benchmark and the counterfactual recession scenarios. λ denotes the aggregate fraction of traders that acquire information. Output and investment values represent the percentage changes from the pre-shock values.

[26]We use 1000 grid points for capital. For deterministic IRFs, we simulate the economy for 220 periods and discard the first 195. For stochastic simulations, we simulate 300 economies for 200 periods and discard the first 100 periods. We present the averages of time series statistics across these economies.

In the case of the simultaneous shock to productivity and liquidity needs (solid lines), the participation of more day traders absconds the fundamental value of firms. As a response to more uncertain fundamentals, the benefits of acquiring information increase for night traders as well. In spite of more information acquisition overall, price informativeness ultimately declines. The decline in aggregate productivity and the increased misallocation induced by lower price informativeness discourages investments and magnifies the decline in output and investment.

How about a recession without financial distress? This case of a single shock to productivity (dotted lines) also generates a decline in output and investment, but only half that when accompanied by financial distress. Even though liquidity needs do not directly affect aggregates, they do so indirectly through stock price informativeness. As in the previous case, the increase in uncertainty σ_{θ_d} induces day traders to produce more information, but the higher σ_{ϵ_d} makes them trade less aggressively on their information. This combination generates less 'noise' in markets and makes stock prices more informative about fundamentals, improving the allocation of capital, which partially compensates for the negative productivity shock.

Our result of a recession without financial distress is reminiscent of "cleansing recessions," but for different reasons. In that literature, recessions reduce the cost of reallocating resources to more productive activities. In our case, recessions increase stock market informativeness and allow for better resource allocation. Our work, however, highlights that "cleansing recessions" can turn into "sullying recessions" when they are accompanied by financial distress and heightened liquidity concerns.

5.2 Alternative Information Structures

In the previous exercise, we analyzed the quantitative response of price informativeness to productivity and liquidity shocks and their aggregate consequences for our calibrated economy. Here, we compare these reactions with those of alternative economies in terms of their informational characteristics. The first column of Table 7 shows the changes in aggregate information acquisition ($\lambda = \gamma\lambda_d + (1-\gamma)\lambda_n$), price informativeness, output, and investment when the economy suffers the dual shock on productivity and liquidity needs. This corresponds to the transition from state 3 to state 1 illustrated by solid curves in Figure 6. Changes are expressed in the first column of Table 7 as percentages relative to "normal times". This is a recession with financial distress reduces price informativeness by 10% and output by almost 15%.

One alternative economy assumes that a fraction of traders receive signals exogenously. This is captured in the second column, which assumes λ does not change when the economy suffers a recession with financial distress.[27] This economy would suffer a very large reduction of price informativeness of 63% (six times larger than the benchmark with endogenous information), which leads to severe misallocation that reduces investment and output by more than 30%. This result highlights the quantitative relevance of endogenizing information acquisition. In the economy, agents react to lower information content in stock markets by acquiring information and partially offsetting such reduction. Information acquisition provides a stabilizing effect to the dual shock on productivity and liquidity needs.

Another alternative economy we consider is one with lower information costs. Since there are two types of traders who acquire different information in our setting, we consider two situations, shown in the last two columns of Table 7, with half the cost of information for night and day traders, respectively. If night traders can acquire information at half the cost (third column), they would acquire more information, price informativeness would decline only 6.5%, leading output to decline 12.5% instead of 14.6%, and investment to decline 13% instead of 18%. In contrast, if day traders can acquire information at half the cost (last column), they would also acquire more information, price informativeness would decline by 11.7% instead of 10%, leading output to decline 15.5% instead of 14.6%, and investment to decline 19% instead of 18%.

An economy with lower information costs for night traders, ν_n, corresponds for instance to improvements on the Securities and Exchange Commission disclosure regulations, with more information requirements to be filed with each prospectus, more frequent filings, or firm-specific statistics publications. In contrast, lower information costs for day traders, ν_d, corresponds to regulations that disclose information about market transactions, such as recent changes that disclose stress tests, the use of Central Counterparties (CCPs), disclosures about trading positions or use of discount windows. Our analysis shows that regulations that facilitate information about firms' profitability make

[27] This fixed level is the average λ value that arises in the boom periods of the simulated benchmark economy.

Table 7: Counterfactual Impulse Responses

Moments	Baseline	Fixed γ	Fixed λ	low ν_n	low ν_d
$\Delta\lambda_{\bar{z}\underline{\gamma}\rightarrow\underline{z}\bar{\gamma}}$	0.167	0.010	0	0.127	0.158
$\Delta PI_{\bar{z}\underline{\gamma}\rightarrow\underline{z}\bar{\gamma}}$	-0.099	0.019	-0.628	-0.065	-0.117
$\Delta Y_{\bar{z}\underline{\gamma}\rightarrow\underline{z}\bar{\gamma}}$	-0.146	-0.084	-0.343	-0.125	-0.155
$\Delta Inv_{\bar{z}\underline{\gamma}\rightarrow\underline{z}\bar{\gamma}}$	-0.180	-0.090	-0.381	-0.130	-0.190

Notes: Each number denotes the percentage change in the moment when the economy moves from a long sequence of $\bar{z}\underline{\gamma}$ states to a $\underline{z}\bar{\gamma}$ state. For the fixed γ scenario, this is equivalent to moving to $\underline{z}\bar{\gamma}$.

Table 8: Counterfactual Stochastic Steady State Averages

Moments	Baseline	Fixed γ	Fixed λ	low ν_n	low ν_d
Y	0.168	0.208	0.129	0.186	0.159
C	0.066	0.082	0.041	0.074	0.062
Inv	0.102	0.125	0.088	0.112	0.096
R	0.122	0.125	0.103	0.124	0.122
PI	0.868	0.948	0.600	0.913	0.849
λ_n	0.288	0.248	0.150	0.372	0.309
λ_d	0.065	0.068	0.086	0.065	0.075

Notes: Each number denotes the percentage change in the moment when the economy moves from a long sequence of $\bar{z}\underline{\gamma}$ states to a $\underline{z}\bar{\gamma}$ state. For the fixed γ scenario, this is equivalent to moving to $\underline{z}\bar{\gamma}$. R denotes the gross interest rate that would clear the market in a competitive economy.

the economy more resilient to recessions with financial shocks, but those that facilitate information about markets' operations do the opposite.

While Table 7 shows how alternative economies fare when facing both productivity and liquidity shocks *on impact*, Table 8 shows how the aggregates from these alternative economies would differ in their stochastic steady state. For instance, economies with lower information costs would display more information acquisition by all traders overall. However, if information about fundamentals is cheaper, the economy displays higher levels of investment, consumption, and output in a stochastic steady state, with the opposite happening if information about markets' operations and assets' liquidity is cheaper.

6 Conclusion

Recessions are characterized by decreased investment and productivity, increased misallocation of capital, and information production activities. We suggest an increase in the concern for asset liquidity may be behind all these patterns. We build a model of stock trading with costly information acquisition where noise in prices is endogenously determined. Then, we introduce this stock trading model into a heterogeneous-firm real business cycle model where the real sector observes the stock markets to learn about investment opportunities. The model can simultaneously generate an increase in information acquisition and an increase in input misallocation by a shock to the number of traders that value asset liquidity. It also allows separating the decline in output due to costs of information acquisition and increased misallocation of capital across firms.

The model provides a mapping between the evolution of structural parameters and stock prices' behavior over time. Through this mapping, we estimate the stock price informativeness for several countries and confirm that it declines in periods of heightened liquidity distress. Our estimated model suggests that declining price informativeness during such periods leads to quantitatively important declines in productivity: the output loss associated with a recession almost doubles when the recession is accompanied by financial distress. Our counterfactuals suggest facilitating easier access to information about firms' fundamentals would lead to higher levels of output and consumption with reduced fluctuations while facilitating easier access to liquidity-relevant information does the opposite.

References

Arellano, Cristina, Yan Bai, and Patrick J Kehoe (2019), "Financial frictions and fluctuations in volatility," *Journal of Political Economy*, 127, 2049–2103.

Bai, Jennie, Thomas Philippon, and Alexi Savov (2016), "Have financial markets become more informative?" *Journal of Financial Economics*, 122, 625–654.

Barlevy, Gadi and Pietro Veronesi (2000), "Information acquisition in financial markets," *The Review of Economic Studies*, 67, 79–90.

Baron, Matthew, Emil Verner, and Wei Xiong (2021), "Banking crises without panics," *The Quarterly Journal of Economics*, 136, 51–113.

Benhabib, Jess, Xuewen Liu, and Pengfei Wang (2019), "Financial markets, the real economy, and self-fulfilling uncertainties," *The Journal of Finance*, 74, 1503–1557.

Bloom, Nicholas, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J Terry (2018), "Really uncertain business cycles," *Econometrica*, 86, 1031–1065.

Breon-Drish, Bradyn (2015), "On existence and uniqueness of equilibrium in a class of noisy rational expectations models," *The Review of Economic Studies*, 82, 868–921.

Christiano, Lawrence J, Roberto Motto, and Massimo Rostagno (2014), "Risk shocks," *American Economic Review*, 104, 27–65.

Cooper, Russell W and Immo Schott (2023), "Capital reallocation and the cyclicality of aggregate productivity," *Quantitative Economics*, 14, 1337–1365.

David, Joel M, Hugo A Hopenhayn, and Venky Venkateswaran (2016), "Information, misallocation, and aggregate productivity," *The Quarterly Journal of Economics*, 131, 943–1005.

David, Joel M and Venky Venkateswaran (2019), "The sources of capital misallocation," *American Economic Review*, 109, 2531–67.

Dávila, Eduardo and Cecilia Parlatore (2018), "Identifying price informativeness," Tech. rep., National Bureau of Economic Research.

Di Nola, Alessandro (2016), "Capital Misallocation During the Great Recession," Capital Misallocation During the Great Recession (August 1, 2016). Available at SSRN: https://ssrn.com/abstract=3033789 or http://dx.doi.org/10.2139/ssrn.3033789.

Dias, Daniel A., Carlos Robalo Marques, and Christine Richmond (2016), "Misallocation and productivity in the lead up to the Eurozone crisis," *Journal of Macroeconomics*, 49, 46 – 70.

Eisfeldt, Andrea L. and Adriano A. Rampini (2006), "Capital reallocation and liquidity," *Journal of Monetary Economics*, 53, 369 – 399.

Eisfeldt, Andrea L. and Adriano A. Rampini (2008), "Managerial incentives, capital reallocation, and the business cycle," *Journal of Financial Economics*, 87, 177 – 199.

Fajgelbaum, Pablo D, Edouard Schaal, and Mathieu Taschereau-Dumouchel (2017), "Uncertainty traps," *The Quarterly Journal of Economics*, 132, 1641–1692.

Farboodi, Maryam and Laura Veldkamp (2020), "Long-run growth of financial data technology," *American Economic Review*, 110, 2485–2523.

Foster, Lucia, Cheryl Grim, and John Haltiwanger (2016), "Reallocation in the great recession: cleansing or not?" *Journal of Labor Economics*, 34, S293–S331.

Fuchs, William, Brett Green, and Dimitris Papanikolaou (2016), "Adverse selection, slow-moving capital, and misallocation," *Journal of Financial Economics*, 120, 286 – 308.

Garman, Mark B and Michael J Klass (1980), "On the estimation of security price volatilities from historical data," *Journal of business*, 67–78.

Goldstein, Itay, Yan Li, and Liyan Yang (2014), "Speculation and hedging in segmented markets," *The Review of Financial Studies*, 27, 881–922.

Gospodinov, Nikolay and Damba Lkhagvasuren (2014), "A moment-matching method for approximating vector autoregressive processes by finite-state Markov chains," *Journal of Applied Econometrics*, 29, 843–859.

Grossman, Sanford J and Joseph E Stiglitz (1980), "On the impossibility of informationally efficient markets," *The American economic review*, 70, 393–408.

Hassan, Tarek A and Thomas M Mertens (2017), "The social cost of near-rational investment," *American Economic Review*, 107, 1059–1103.

Kehrig, Matthias (2015), "The Cyclical Nature of the Productivity Distribution," Available at SSRN: https://ssrn.com/abstract=1854401.

Khan, Aubhik and Julia K Thomas (2013), "Credit shocks and aggregate fluctuations in an economy with production heterogeneity," *Journal of Political Economy*, 121, 1055–1107.

Oberfield, Ezra (2013), "Productivity and misallocation during a crisis: Evidence from the Chilean crisis of 1982," *Review of Economic Dynamics*, 16, 100 – 119, special issue: Misallocation and Productivity.

Ordonez, Guillermo (2013), "The asymmetric effects of financial frictions," *Journal of Political Economy*, 121, 844–895.

Reuters, Thomson (2010), "Worldscope Database Data definitions guide," *NY: Thomson Reuters*.

Sandleris, Guido and Mark L. J. Wright (2014), "The Costs of Financial Crises: Resource Misallocation, Productivity, and Welfare in the 2001 Argentine Crisis," *The Scandinavian Journal of Economics*, 116, 87–127.

Sockin, Michael (2015), "Not so great expectations: A model of growth and informational frictions," *mimeo*.

Stein, Jeremy C (1987), "Informational externalities and welfare-reducing speculation," *Journal of political economy*, 95, 1123–1145.

Straub, Ludwig and Robert Ulbricht (2023), "Endogenous Uncertainty and Credit Crunches," *Review of Economic Studies*, page rdad110.

Vanden, Joel M (2008), "Information quality and options," *The Review of Financial Studies*, 21, 2635–2676.

Vives, Xavier (2014), "On the possibility of informationally efficient markets," *Journal of the European Economic Association*, 12, 1200–1239.

Wang, Jiang (1993), "A model of intertemporal asset prices under asymmetric information," *The Review of Economic Studies*, 60, 249–282.

A Proofs

Proof of Lemma 1. Conjecture a linear price function for each aggregate state s:

$$p_i^s = \phi_{i0}^s + \phi_{in}^s \theta_{in} + \phi_{id}^s \theta_{id} + \phi_{i\epsilon}^s \tilde{\epsilon}_{in}^- \tag{27}$$

Then, the signals uninformed traders will use from observing the price can be drawn from $(p_i^s - \phi_{i0}^s - \phi_{i\epsilon}^s \tilde{\epsilon}_{in}^- - \phi_{id}^s \theta_{id})/\phi_{in}$ and $(p_i^s - \phi_{i0}^s - \phi_{i\epsilon}^s \tilde{\epsilon}_{in}^- - \phi_{in}^s \theta_{in})/\phi_{id}$ for θ_{in} and θ_{id}, respectively. Since the prior distributions are Gaussian and the signal is a linear function of a Gaussian random variable, the posterior distribution for z_{in} is also Gaussian with mean and variance:

$$E_s\left[z_{in} \mid p_i\right] = Z + \frac{\frac{\bar{\theta}_{in}}{\sigma_{\theta_{in}}^2} + \frac{1}{\sigma_{\theta_{id}}^2}\left(\frac{\phi_{in}^s}{\phi_{id}^s}\right)^2 \left(\frac{p_i - \phi_{i0}^s - \phi_{id}^s \bar{\theta}_{id} - \phi_{i\epsilon}^s \tilde{\epsilon}_{in}^-}{\phi_{in}^s}\right)}{\frac{1}{\sigma_{\theta_{in}}^2} + \frac{1}{\sigma_{\theta_{id}}^2}\left(\frac{\phi_{in}^s}{\phi_{id}^s}\right)^2} + \rho \tilde{\epsilon}_{in}^- \tag{28}$$

$$\mathrm{Var}_s\left[z_{in} \mid p_i\right] = \left(\frac{1}{\sigma_{\theta_{in}}^2} + \frac{1}{\sigma_{\theta_{id}}^2}\left(\frac{\phi_{in}^s}{\phi_{id}^s}\right)^2\right)^{-1} + \sigma_{\epsilon_{in}}^2 \tag{29}$$

Using these, we can write down the expectation and the variance for the total payoff from holding one share of the firm i:

$$E_s\left[z_{in} + p_i' \mid p_i\right] = E_s\left[z_{in} \mid p_i\right] + \sum_{s'} q_{ss'}[\phi_{io}^{s'} + \phi_{in}^{s'} \bar{\theta}_{in} + \phi_{id}^{s'} \bar{\theta}_{id} + \phi_{i\epsilon}^{s'} \rho \tilde{\epsilon}_{in}^-] \tag{30}$$

$$\begin{aligned} \mathrm{Var}_s\left[z_{in} + p_i' \mid p_i\right] &= \mathrm{Var}_s\left(\theta_{in} \mid p_i\right) + \mathrm{Var}_s\left(\epsilon_{in} + p_i'\right) \\ &= \left(\frac{1}{\sigma_{\theta_{in}}^2} + \frac{1}{\sigma_{\theta_{id}}^2}\left(\frac{\phi_{in}^s}{\phi_{id}^s}\right)^2\right)^{-1} \\ &\quad + \mathrm{Var}_s\left(\phi_{0i}^{s'} + \phi_{di}^{s'} \theta_{di} + \phi_{ni}^{s\prime} \theta_{ni} + \left(1 + \phi_{\epsilon i}^{s'} \rho^2\right) \epsilon_{in} + \phi_{\epsilon i}^{s'} \rho \tilde{\epsilon}_{in}^-\right) \end{aligned} \tag{31}$$

Using these, we can rewrite the market clearing condition as

$$(1-\gamma)\lambda_{in}^{s}\left[\frac{Z+\tilde{\theta}_{in}+E_{s}\left[p_{i}'\right]-(1+r^{I})p_{i}}{a\,\mathrm{Var}_{s}\left(\epsilon_{in}+p_{i}'\right)}\right]+\gamma\lambda_{id}^{s}\left[\frac{Z+\tilde{\theta}_{in}-\theta_{id}+E_{s}[p_{i}']-(1+r^{I})p_{i}}{a\left(\sigma_{\epsilon_{id}}^{2}+\mathrm{Var}_{s}\left(\epsilon_{in}+p_{i}'\right)\right)}\right]+$$
$$(1-\gamma)\left(1-\lambda_{in}^{s}\right)\left[\frac{E_{s}\left[z_{in}\mid p_{i}\right]+E_{s}\left[p_{i}'\right]-(1+r^{I})p_{i}}{a\left(\mathrm{Var}_{s}\left(\epsilon_{in}+p_{i}'\right)+\mathrm{Var}_{s}\left(\theta_{in}\mid p_{i}\right)\right)}\right]+$$
$$\gamma\left(1-\lambda_{id}^{s}\right)\left[\frac{E_{s}\left[z_{in}-z_{id}\mid p_{i}\right]+E_{s}\left[p_{i}'\right]-(1+r^{I})p_{i}}{a\left(\mathrm{Var}_{s}\left(\epsilon_{in}+p_{i}'\right)+\mathrm{Var}_{s}\left(\theta_{in}-\theta_{id}\mid p_{i}\right)\right)}\right]=\overline{K_{i}} \tag{32}$$

We suppress the aggregate state s in the rest of the proof to declutter the notation. First, denote

$$\chi_{1}=\frac{\gamma\lambda_{id}}{a\left(\sigma_{\epsilon_{id}}^{2}+\mathrm{Var}\left(\epsilon_{in}+p_{i}'\right)\right)} \qquad \chi_{3}=\frac{\gamma\left(1-\lambda_{id}\right)}{a\left(\sigma_{\epsilon_{id}}^{2}+\mathrm{Var}\left(\epsilon_{in}+p_{i}'\right)+\mathrm{Var}\left(\theta_{in}-\theta_{id}\mid p_{i}\right)\right)}$$
$$\chi_{2}=\frac{(1-\gamma)\lambda_{in}}{a\,\mathrm{Var}\left(\epsilon_{in}+p_{i}'\right)}, \qquad \chi_{4}=\frac{(1-\gamma)\left(1-\lambda_{in}\right)}{a\left(\mathrm{Var}\left(\epsilon_{in}+p_{i}'\right)+\mathrm{Var}\left(\theta_{in}\mid p_{i}\right)\right)} \tag{33}$$

and $\chi=(\chi_{1}+\chi_{2}+\chi_{3}+\chi_{4})$. One can rearrange the terms to get

$$\left(\chi_{1}+\chi_{2}\right)\left(\theta_{in}+\rho\tilde{\epsilon}_{in}^{-}\right)-\chi_{1}\theta_{id}+\chi\left(Z+\sum_{s'}\left[\phi_{i0}^{s'}+\phi_{in}^{s'}\bar{\theta}_{n}+\phi_{id}^{s'}\bar{\theta}_{d}+\phi_{i\epsilon}^{s'}\rho\tilde{\epsilon}_{in}^{-}\right]q_{ss'}\right)$$
$$+\left(\chi_{3}+\chi_{4}\right)\left(\rho\tilde{\epsilon}_{in}^{-}+\frac{\frac{\bar{\theta}_{in}}{\sigma_{\theta_{in}}^{2}}+\frac{1}{\sigma_{\theta_{id}}^{2}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^{2}\left(\frac{p_{i}-\phi_{i0}-\phi_{id}\bar{\theta}_{id}-\phi_{i\epsilon}\tilde{\epsilon}_{in}^{-}}{\phi_{in}}\right)}{\frac{1}{\sigma_{\theta_{in}}^{2}}+\frac{1}{\sigma_{\theta id}^{2}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^{2}}\right)-\chi_{3}\frac{\frac{\bar{\theta}_{id}}{\sigma_{\theta_{id}}^{2}}+\frac{1}{\sigma_{\theta_{in}}}\left(\frac{\phi_{id}}{\phi_{in}}\right)^{2}\left(\frac{p_{i}-\phi_{i0}-\phi_{in}\bar{\theta}_{in}-\phi_{i\epsilon}\tilde{\epsilon}_{in}^{-}}{\phi_{id}}\right)}{\frac{1}{\sigma_{\theta_{id}}^{2}}+\frac{1}{\sigma_{\theta_{in}}^{2}}\left(\frac{\phi_{id}}{\phi_{in}}\right)^{2}}$$
$$-\chi(1+r^{I})p_{i}=\bar{K}_{i} \tag{34}$$

Next, we rearrange terms to leave p_i alone:

$$\underbrace{\left[(1+r^I)\chi+\chi_3\frac{\left(\frac{\phi_{id}}{\phi_{in}}\right)^2\frac{1}{\sigma^2_{\theta_{in}}}\operatorname{Var}\left[\theta_{id}\mid p_i\right]}{\phi_{id}}-(\chi_3+\chi_4)\frac{\left(\frac{\phi_{in}}{\phi_{id}}\right)^2\frac{1}{\sigma_{\theta_{id}}}\operatorname{Var}\left[\theta_{in}\mid p_i\right]}{\phi_{in}}\right]}_{\tilde{\phi}}p_i=\underbrace{(\chi_1+\chi_2)}_{\tilde{\phi}^s_{in}}\theta_{in}$$

$$-\underbrace{\chi_1}_{\tilde{\phi}^s_{id}}\theta_{id}+\left[\rho\chi\left(1+\sum_s q_{ss'}\phi^{s'}_{i\epsilon}\right)-(\chi_3+\chi_4)\left(\frac{\operatorname{Var}\left[\theta_{in}\mid p_i\right]}{\sigma^2_{\theta id}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2\frac{\phi_{i\epsilon}}{\phi_{in}}\right)\right.$$

$$\left.+\chi_3\frac{\operatorname{Var}\left[\theta_{id}\mid p_i\right]}{\sigma^2_{\theta in}}\left(\frac{\phi_{id}}{\phi_{in}}\right)^2\frac{\phi_{i\epsilon}}{\phi_{id}}\right]\tilde{\epsilon}^-_{in}+\chi\left(Z+\sum_{s'}\left[\phi^{s'}_{i0}+\phi^{s'}_{in}\bar{\theta}_{in}+\phi^{s'}_{id}\bar{\theta}_{id}\right]q_{ss'}\right)+$$

$$+\chi_3\operatorname{Var}\left[\theta_{id}\mid p_i\right]\left[\frac{\phi_{0i}+\phi_{ni}\bar{\theta}_{in}}{\phi_{di}}\left(\frac{\phi_{id}}{\phi_{in}}\right)^2\frac{1}{\sigma^2_{\theta_{in}}}-\frac{\bar{\theta}_{id}}{\sigma^2_{\theta_{id}}}\right]-\bar{K}_i$$

$$-(\chi_3+\chi_4)\operatorname{Var}\left[\theta_{in}\mid p_i\right]\left[\frac{\phi^s_{0i}+\phi^s_{id}\bar{\theta}_{id}}{\phi_{in}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2\frac{1}{\sigma^2_{\theta_{id}}}-\frac{\bar{\theta}_{in}}{\sigma^2_{\theta_{in}}}\right] \tag{35}$$

In (35), $\phi^s_{in}=\frac{\tilde{\phi}^s_{in}}{\tilde{\phi}}$ and $\phi^s_{id}=\frac{\tilde{\phi}^s_{id}}{\tilde{\phi}}$. Hence the expression in the Lemma follows. □

Proof of Lemma 1. The proof extends the corresponding proof in Grossman and Stiglitz (1980) to this environment. Since the end-of-period wealth is additive in the information acquisition costs and payoffs across stocks, proving the result for a single-asset case is sufficient. First, notice that the end-of-period wealth for informed and uninformed agents can be written as

$$\begin{aligned}W^{n,i}_{I,j}&=r\left(W_{oj}-c\left(\lambda^i_n\right)\right)+[(z_{in}+p'_i)-(1+r^I)p_i]X^n_{iI}\\W^{n,i}_{U,j}&=rW_{oj}+[(z_{in}+p'_i)-(1+r^I)p_i]X^n_{iU}\end{aligned}$$

The expected value of being informed for a night trader j can be written as

$$E[V(W^{n,i}_{I,j})\mid p]=E[e^{-aW^{n,i}_{I,j}}\mid p_i]=-exp\left(-aE\left[E[W^{n,i}_{I,j}\mid\theta]-\frac{a}{2}\operatorname{Var}\left[W^{n,i}_{I,j}\mid\theta\right]\Big|p_i\right]\right) \tag{36}$$

Combining the three equations, we can write

$$E[W_{I,j}^{n,i} \mid \theta] = r\left(W_{oj} - c\left(\lambda_n^i\right)\right) + \frac{[Z + \tilde{\theta}_{in} + E[p_i'] - (1 + r^I)p_i]^2}{a\,\mathrm{Var}(z_{in} + p_i' \mid \theta)} \tag{37}$$

$$\mathrm{Var}[W_{I,j}^{n,i} \mid \theta] = \frac{[Z + \tilde{\theta}_{in} + E[p_i'] - (1 + r^I)p_i]^2}{a^2\,\mathrm{Var}(z_{in} + p_i' \mid \theta)} = \frac{E[W_{I,j}^{n,i} \mid \theta]}{a} \tag{38}$$

since W_{0j} and p_i are not random given θ. Thus, we can rewrite Equation 36 as

$$\begin{aligned} E[V(W_{I,j}^{n,i}) \mid p] &= -\,exp\left[ar(W_{oj} - c(\lambda_n^i)) - \frac{[Z + \tilde{\theta}_{in} + E[p_i'] - (1 + r^I)p_i]^2}{2\,\mathrm{Var}(z_{in} + p_i' \mid \theta)}\right] \\ &= -\,exp\left[-ar(W_{oj} - c(\lambda_n^i))\right] \times \\ &\quad E\left[exp\left(\frac{-1}{2\,\mathrm{Var}(z_{in} + p_i' \mid \theta)}\left[Z + \tilde{\theta}_{in} + E[p_i'] - (1 + r^I)p_i\right]^2\right) \mid p_i\right] \end{aligned} \tag{39}$$

Now define

$$\begin{aligned} h_{in} &:= \mathrm{Var}[\tilde{\theta}_{in} \mid p] \\ g_{in} &:= \frac{Z + \tilde{\theta}_{in} + E[p_i'] - (1 + r^I)p_i}{\sqrt{h_{in}}} \end{aligned}$$

so, $E[V(W_{I,j}^{n,i}) \mid p]$ can be rewritten as

$$E[V(W_{I,j}^{n,i}) \mid p] = e^{ac^n(X_{in})}V(rW_{oj})E_s\left[exp\left(\frac{-h_{in}}{\mathrm{Var}(z_{in} + p_i' \mid \theta)}g_{in}^2\right) \mid p_i\right] \tag{40}$$

Since p_i is a linear function of θ, conditional on p_i, $\tilde{\theta}_{in}$ is normally distributed. Therefore, g_{in}^2 is distributed with Chi-squared. Hence, moment generating function of g_{in}^2 has the form:[28]

[28]For this to work, we need $\mathrm{Var}(z_{in} + p_i' \mid \theta)$ to be deterministic given p_i, i.e., $\mathrm{Var}[\mathrm{Var}(z_{in} + p_i' \mid \theta) \mid p_i] = 0$. $\mathrm{Var}(z_{in} + p_i' \mid \theta) = \mathrm{Var}(\epsilon_{in} + p_i')$ is not a function of θ or a function of p_i.

$$E[e^{-t(g_{in})^2}|p] = \frac{1}{\sqrt{1+2t}} exp\left(\frac{-t(E[g_{in} \mid p])^2}{1+2t}\right). \tag{41}$$

Now we can rewrite,

$$\begin{aligned} E[V(W_{I,j}^{n,i}) \mid p] =& \frac{1}{\sqrt{1+\frac{h_{in}}{\mathrm{Var}(z_{in}+p'_i|\theta)}}} exp\left(\frac{-h_{in}E[g_{in} \mid p]^2}{2\left(\mathrm{Var}(z_{in}+p'_i \mid \theta)+h_{in}\right)}\right) \\ =& \frac{1}{\sqrt{1+\frac{h_{in}}{\mathrm{Var}(z_{in}+p'_i|\theta)}}} exp\left(\frac{-\left(Z+E[\tilde{\theta}_{in} \mid p_i]+E[p'_i]-(1+r^I)p_i\right)^2}{2\left(\mathrm{Var}(z_{in}+p'_i \mid \theta)+h_{in}\right)}\right) \end{aligned} \tag{42}$$

Furthermore, notice that

$$\frac{\mathrm{Var}\left(z_{in}+p'_i \mid \theta\right)}{\mathrm{Var}\left(z_{in}+p'_i \mid p_i\right)} = \frac{\mathrm{Var}\left(z_{in}+p'_i \mid \theta\right)}{\mathrm{Var}\left(z_{in}+p'_i \mid \theta\right)+h_{in}} = \frac{1}{1+\frac{h_{in}}{\mathrm{Var}\left(z_{in}+p'_i|\theta\right)}} \tag{43}$$

Hence, we can rewrite as

$$\begin{aligned} E\left[\exp\left(\frac{-h_{in}}{\mathrm{Var}\left(z_{in}+p'_i \mid \theta\right)} g_{in}^2\right)\middle| p_i\right] =& \sqrt{\frac{\mathrm{Var}\left(z_{in}+p'_i \mid \theta\right)}{\mathrm{Var}\left(z_{in}+p'_i \mid p_i\right)}} \times \\ & \exp\left(\frac{-\left(Z+E\left[\tilde{\theta}_{in} \mid p_i\right]+E\left[p'_i\right]-(1+r^I)p_i\right)^2}{2\left(\mathrm{Var}\left(z_{in}+p'_i \mid \theta\right)+\mathrm{Var}\left[\theta_{in} \mid p_i\right]\right)}\right) \end{aligned} \tag{44}$$

Then,

$$E\left[V(W_I^{n,i}j) \mid p\right] = e^{ac(\lambda_{in})} V(rW_{oj}) \sqrt{\frac{\mathrm{Var}\left(z_{in} + p_i' \mid \theta\right)}{\mathrm{Var}\left(z_{in} + p_i' \mid p_i\right)}} \times \exp\left(\frac{-\left(Z + E\left[\tilde{\theta}_{in} \mid p_i\right] + E\left[p_i'\right] - (1 + r^I)p_i\right)^2}{2\left(\mathrm{Var}\left(z_{in} + p_i' \mid \theta\right) + \mathrm{Var}\left[\theta_{in} \mid p_i\right]\right)} \right) \tag{45}$$

Following similar steps for the value of being uninformed yields

$$E\left[V\left(W_{u,j}^{n,i}\right) \mid P\right] = V\left(rW_{o_j}\right) \exp\left(\frac{-\left(Z + E\left[\tilde{\theta}_{in} \mid p_i\right] + E\left[p_i'\right] - (1 + r^I)p_i\right)^2}{2\left(\mathrm{Var}\left(z_{in} + p_i' \mid \theta\right) + \mathrm{Var}\left[\theta_{in} \mid p_i\right]\right)} \right) \tag{46}$$

Therefore,

$$\frac{E\left[V\left(W_{I,j}^{n,i}\right) \mid P\right]}{E\left[V\left(W_{u,j}^{n,i}\right) \mid P\right]} = e^{ac(\lambda_{in})} \sqrt{\frac{\mathrm{Var}\left(z_{in} + p_i' \mid \theta\right)}{\mathrm{Var}\left(z_{in} + p_i' \mid \theta\right) + \mathrm{Var}\left(\theta_{in} \mid p_i\right)}}$$

□

Proof of Proposition 2. The ex-ante bias associated with the hedge fund's estimator can be written as:

$$\begin{aligned}
\left|E\left[E[\theta_{in} \mid p_i] - \theta_{in}\right]\right| &= \left| \frac{\frac{\bar{\theta}_{in}}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2 \left(\frac{p_i - \phi_{i0} - \phi_{id}\bar{\theta}_{id} - \phi_{i\epsilon}\bar{\tilde{\epsilon}}_{in}}{\phi_{in}}\right)}{\frac{1}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2} - \bar{\theta}_{in} \right| \\
&= \left| \frac{\frac{\bar{\theta}_{in}}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2 \left(\frac{\phi_{i0} + \phi_{id}\bar{\theta}_{id} + \phi_{in}\bar{\theta}_{in} + \phi_{i\epsilon}\bar{\tilde{\epsilon}}_{in} - \phi_{i0} - \phi_{id}\bar{\theta}_{id} - \phi_{i\epsilon}\bar{\tilde{\epsilon}}_{in}}{\phi_{in}}\right)}{\frac{1}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2} - \bar{\theta}_{in} \right| \\
&= \left| \frac{\frac{\bar{\theta}_{in}}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2 \bar{\theta}_{in}}{\frac{1}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2} - \bar{\theta}_{in} \right| = 0.
\end{aligned}$$

The variance associated with the estimator can be written as:

$$\begin{aligned}\text{Var}\left[E[\theta_{in} \mid p_i] - \theta_{in}\right] =& \text{Var}\left[\frac{\frac{\bar{\theta}_{in}}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2\left(\frac{\phi_{id}(\theta_{id}-\bar{\theta}_{id})+\phi_{in}\theta_{in}}{\phi_{in}}\right)}{\frac{1}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2} - \theta_{in}\right] \\ =& \frac{\text{Var}\left[\frac{\bar{\theta}_{in}}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2\left(\frac{\phi_{id}\theta_{id}}{\phi_{in}}\right) - \frac{\theta_{in}}{\sigma^2_{\theta_{in}}}\right]}{\left(\frac{1}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2\right)^2} \\ =& \frac{\frac{1}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2}{\left(\frac{1}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2\right)^2} = \frac{1}{\frac{1}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2}.\end{aligned}$$

Lastly, under the mean-squared error loss function, $R\left(\theta_{in}, E[\theta_{in}|p_i]\right) = \left|E\left[E[\theta_{in} \mid p_i] - \theta_{in}\right]\right|^2 + \text{Var}\left[E[\theta_{in} \mid p_i] - \theta_{in}\right]$. Hence

$$R\left(\theta_{in}, E[\theta_{in}|p_i]\right) = 0 + \frac{1}{\frac{1}{\sigma^2_{\theta_{in}}} + \frac{1}{\sigma^2_{\theta_{id}}}\left(\frac{\phi_{in}}{\phi_{id}}\right)^2}. \tag{47}$$

□

Proof of Proposition 3. The pricing function under $\rho = 0$ becomes

$$\begin{aligned}P_i =& \Phi_o + \Phi_n\theta_{in} + \Phi_d\theta_{id} \\ =& \Phi_o + \Phi_n z_{in} + \Phi_d z_{id} - \Phi_n\epsilon_{in} - \Phi_d\epsilon_{id}.\end{aligned}$$

Hence, when the price is regressed on θ_{in} and θ_{id}, the error term becomes $\nu_i = -\Phi_n\epsilon_{in} - \Phi_d\epsilon_{id}$, which is correlated with z_{in}. Let $\tilde{Z}_i$=$[1 \quad z_{in} \quad z_{id}]$,$\tilde{\Phi} = [\Phi_o \quad \Phi_n \quad \Phi_d]$. Then

$$\hat{\Phi}_{OLS} = \tilde{\Phi} - \Phi_n\left(\frac{1}{n}\sum \tilde{Z}_i\tilde{Z}_i'\right)^{-1}\left(\frac{1}{n}\sum \tilde{Z}_i\epsilon_{in}\right) - \Phi_d\left(\frac{1}{n}\sum \tilde{Z}_i\tilde{Z}_i'\right)^{-1}\left(\frac{1}{n}\sum \tilde{Z}_i\epsilon_{id}\right).$$

First, because θ_{in}, θ_{id}, and ϵ_{in} are independent, the second term on the right-hand side can be decomposed as:

$$\left(\frac{1}{n}\sum \tilde{Z}_i\tilde{Z}_i'\right)^{-1}\left(\frac{1}{n}\sum \tilde{Z}_i\epsilon_{in}\right) = \left(\frac{1}{n}\sum \tilde{Z}_i Z_i'\right)^{-1}\left(\frac{1}{n}\sum [\epsilon_{in} \quad \theta_{in}\epsilon_{in} + \epsilon_{in}^2 \quad \theta_{id}\epsilon_{in} + \epsilon_{id}\epsilon_{in}]\right)$$

$$\xrightarrow{\text{p}} \underbrace{\begin{bmatrix} 1 & \bar{Z}_n & \bar{Z}_d \\ \bar{Z}_n & \overline{Z_n Z_n} & \bar{Z}_n\bar{Z}_d \\ \bar{Z}_d & \bar{Z}_n\bar{Z}_d & \overline{Z_d Z_d} \end{bmatrix}^{-1}}_{\mathbf{Z}^{-1}} \begin{bmatrix} 0 \\ \sigma_{\epsilon_n}^2 \\ 0 \end{bmatrix},$$

where $\xrightarrow{\text{p}}$ denotes convergence in probability and $\bar{X}_n$ denotes $E[X_i]$. We can further write

$$Z^{-1} = \frac{1}{det(z)}\begin{bmatrix} (\sigma_{\epsilon_n}^2 + \sigma_{\theta_n}^2 + \overline{\theta}_n^2(\sigma_{\epsilon_d}^2 + \sigma_{\theta_d}^2 + \overline{\theta}_d^2) - \overline{\theta}_n^2\overline{\theta}_d^2 & -(\sigma_{\epsilon_d}^2 + \sigma_{\theta_d}^2)\overline{\theta}_n^2 & -\theta_d(\sigma_{\epsilon_n}^2 + \sigma_{\theta_n}^2) \\ -(\sigma_{\epsilon_d}^2 + \sigma_{\theta_d}^2)\overline{\theta}_n & \sigma_{\epsilon_d}^2 + \sigma_{\theta_d}^2 & 0 \\ -\theta_d(\sigma_{\epsilon_n}^2 + \sigma_{\theta_n}^2) & 0 & \sigma_{\theta_n}^2 + \sigma_{\epsilon_n}^2 \end{bmatrix}.$$

Then, we can characterize the term as

$$\left(\frac{1}{n}\sum \tilde{Z}_i\tilde{Z}_i'\right)^{-1}\left(\frac{1}{n}\sum \tilde{Z}_i\epsilon_{in}\right) = Z^{-1}\begin{bmatrix} 0 \\ \sigma_{\epsilon_n}^2 \\ 0 \end{bmatrix} = \frac{1}{det(Z)}\begin{bmatrix} \sigma_{\epsilon_n}^2(\sigma_{\epsilon_d}^2 + \sigma_{\theta_d}^2)\overline{\theta}_n \\ \sigma_{\epsilon_n}^2(\sigma_{\epsilon_d}^2 + \sigma_{\theta_d}^2) \\ 0 \end{bmatrix},$$

where

$$
\begin{aligned}
det(z) =& 1\left((\sigma^2_{\epsilon_n} + \sigma^2_{\theta_n} + \overline{\theta}^2_n)(\sigma^2_{\epsilon_d} + \sigma^2_{\theta_d} + \overline{\theta}^2_d) - \overline{\theta}^2_n\overline{\theta}^2_d\right) - \overline{\theta}_n\left(\overline{\theta}_n(\sigma^2_{\epsilon_d} + \sigma^2_{\theta_d} + \overline{\theta}^2_d) - \overline{\theta}_d\overline{\theta}_n\overline{\theta}_d\right) \\
&+ \overline{\theta}_d\left(\overline{\theta}_n\overline{\theta}_n\overline{\theta}_d - \overline{\theta}_d(\sigma^2_{\epsilon_n} + \sigma^2_{\theta_n} + \overline{\theta}^2_n)\right) \\
=& \sigma^2_{\epsilon_n}\sigma^2_{\epsilon_d} + \sigma^2_{\epsilon_n}\sigma^2_{\theta_d} + \sigma^2_{\epsilon_n}\overline{\theta}^2_d + \sigma^2_{\epsilon_d}\sigma^2_{\theta_n} + \sigma^2_{\epsilon_d}\overline{\theta}^2_n + \sigma^2_{\theta_n}\sigma^2_{\theta_d} + \sigma^2\theta_d\overline{\theta}^2_d + \sigma^2{}_{\theta_d}\overline{\theta}^2_n \\
&- \sigma^2_{\epsilon_d}\sigma^2_{\epsilon_n} - \sigma^2_{\theta_d}\overline{\theta}^2_n - \overline{\theta}^2_d\overline{\theta}^2_n + \overline{\theta}^2_n\overline{\theta}^2_d + \overline{\theta}^2_d\overline{\theta}^2_n - \sigma^2_{\epsilon n}\overline{\theta}^2_d - \sigma^2_{\theta_n}\overline{\theta}^2_d - \overline{\theta}^2_n\overline{\theta}^2_d \\
=& \sigma^2_{\epsilon_n}\sigma^2_{\epsilon_d} + \sigma^2_{\epsilon_n}\sigma^2_{\theta_d} + \sigma^2_{\epsilon_d}\sigma^2_{\theta_n} + \sigma^2_{\theta_n}\sigma^2_{\theta_d}.
\end{aligned}
$$

Following similar steps would yield:

$$
\left(\frac{1}{n}\sum \tilde{Z}_i\tilde{Z}_i'\right)^{-1}\left(\frac{1}{n}\sum \tilde{Z}_i\epsilon_{id}\right) = \frac{1}{det(Z)}\begin{bmatrix}\sigma^2_{\epsilon_d}(\sigma^2_{\epsilon_n} + \sigma^2_{\theta_n})\overline{\theta}_d \\ \\ 0 \\ \\ \sigma^2_{\epsilon_d}(\sigma^2_{\epsilon_n} + \sigma^2_{\theta_n})\end{bmatrix}.
$$

Therefore,

$$
\hat{\Phi}_{OLS} = \widetilde{\Phi} \;-\; \frac{1}{(\sigma^2_{\epsilon_n} + \sigma^2_{\theta_n})(\sigma^2_{\epsilon_d} + \sigma^2_{\theta_d})}\begin{bmatrix}-\sigma^2_{\epsilon_n}(\sigma^2_{\epsilon_d} + \sigma^2_{\theta_d})\overline{\theta}_n\Phi_n - \sigma^2_{\epsilon_d}(\sigma^2_{\epsilon_n} + \sigma^2_{\theta_n})\overline{\theta}_d\Phi_d \\ \\ \sigma^2_{\epsilon_n}(\sigma^2_{\epsilon_d} + \sigma^2_{\theta_d})\Phi_n \\ \\ \sigma^2_{\epsilon_d}(\sigma^2_{\epsilon_n} + \sigma^2_{\theta_n})\Phi_d\end{bmatrix}
$$

$$
= \begin{bmatrix}\Phi_o + \frac{\overline{\theta}_n\sigma^2_{\epsilon_n}\Phi_n}{\sigma^2_{\epsilon_n} + \sigma^2_{\theta_n}} + \frac{\overline{\theta}_d\sigma^2_{\epsilon_d}\Phi_d}{\sigma^2_{\epsilon_d} + \sigma^2_{\theta_d}} \\ \\ \Phi_n(1 - \frac{\sigma^2_{\epsilon_n}}{\sigma^2_{\epsilon_n} + \sigma^2_{\theta_n}}) \\ \\ \Phi_d(1 - \frac{\sigma^2_{\epsilon_d}}{\sigma^2_{\epsilon_d} + \sigma^2_{\theta_d}})\end{bmatrix}.
$$

□

B Data Cleaning

This section describes our methodology for defining an economy (or a market) and the steps we take to standardize data within an economy.

B.1 Market Assignment

There are multiple ways of defining the relevant stock market for a given economy. While both I/B/E/S and Worldscope assign each company to a country, this assignment is not consistent. Worldscope, before 2013, assigned companies based on "... country of major operations revenue of the company and if not determined by operations then country of headquarters", while after 2013, the assignment was based on the primary listing of the company. On the other hand, I/B/E/S assigns companies based on "country of domicile". The assigned country does not always match between the two datasets. To overcome these inconsistencies, we reassign each company to a country based on the location of the stock exchange the company's shares are traded.

First, we remove all the stocks that have multiple nation or industry assignments in Worldscope. Second, we remove stocks that are cross-listed in multiple exchanges.[29] Third, we link stock exchanges to countries using the bridge provided by Worldscope and aggregate the exchanges within a single country. If the stock exchange information is missing, listed as 'others', or the stock is traded over the counter, we assign the stock to a market on the nation variable.

The assignment of companies based on the stock exchange they trade on is not entirely innocuous, even though the disconnect between the country assignment through stock exchanges and base of operations only exists for a small fraction of companies in most markets. Table X provides the fraction of companies that have a mismatch between the country assigned to them by us based on their stock exchange and the country assigned to them by I/B/E/S. Most of the issues concentrate on Chinese companies that trade in

[29] IBES and Worldscope sometimes use different identifiers for the stocks of the same company in different exchanges. For example, we've noticed an instance where "SUEZ" and "SUEZ LYONNAISE DES EAUX" refer to the same company but trade on different markets and have different "IBES tickers" ("SZE" and "@LYE"). While Worldscope retains data for both, IBES only selects and collects the forecast for "@LYE." Unfortunately, it is not possible to systematically deal with these instances.

Hong Kong and Singapore.

B.2 Exchange Rate Adjustments

In this section, we describe how we standardize the exchange rates for each market to allow cross-sectional and time-series comparability.

First, for countries that have adopted the Euro as their exchange rate, we convert all numbers to the country's original currency using the exchange rate at the time of adoption. This allows the use of exchange rates before the Euro was first introduced. Second, for each market, we determine the dominant currency using the most commonly used currency across its stocks in Worldscope. Third, we convert all values in a market (prices, actuals, forecasts, etc.) to the dominant currency using the date of the closest available exchange rate. We validate our steps by comparing a random sample of our adjusted series with other sources that already present the data in the destination exchange rate. Furthermore, after the exchange rate adjustments, the variances of the cross-sectional measures such as the Price-Earnings ratio are reduced dramatically.

C Robustness Checks

C.1 Monthly Variation in Forecasts

In the baseline analysis, we focus on earnings forecasts that are made six months prior to the fiscal year-end date. This ensures that six months have passed since the end of the prior fiscal year; hence, the associated earnings announcements are already made for most firms. Therefore, the prices at that point already carry the information from the previous year's earnings.

If the forecasts change substantially month-to-month, then our results would be sensitive to the timing assumptions made. Here, we show that the month-to-month variation in earnings forecasts is relatively small. We focus on the forecasts of companies in the U.S. and Japan and set the monthly forecast for a stock as the median forecast made within 15

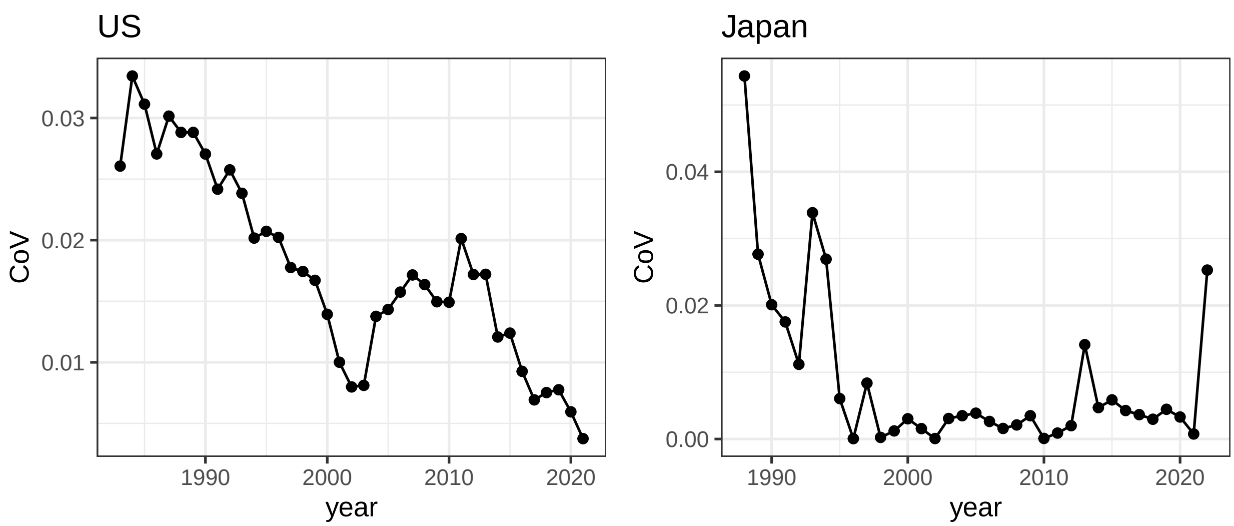


Figure 7: Month-to-Month Coefficient of Variation of Earnings-per-Share Forecasts for the Median Firm Notes:

days of the beginning of the month. We restrict attention to firms whose fiscal years end in the usual months -January in the U.S. and March in Japan- and to forecasts made 4 to 8 months before the fiscal year-end. We only include firms whose forecasts are announced in all months.

Figure 7 shows the monthly variation in forecasts for the median stock. In the US, the monthly standard deviation for the median stock is almost always below 3% of the mean, while it's below 4% for Japan. Hence, the month-to-month variation in forecasts is relatively small around the time we look at the forecasts.

Figure 8 plots the normalized forecast error, i.e., the absolute forecast error divided by the realized value, for the median stock. For both the U.S. and Japan, the forecast error declines as the forecast date becomes closer to the fiscal year's end, with few exceptions. However, the improvement for the median stock mostly stays small. The forecast gets about 5 p.p. and 4 p.p. better in four months for Japan and the U.S., respectively.

C.2 Range Volatility Measures

In this section, we provide summary statistics on the range volatility measure we use. Figure 9 depicts the median firm's normalized range volatility measures for Japan and the

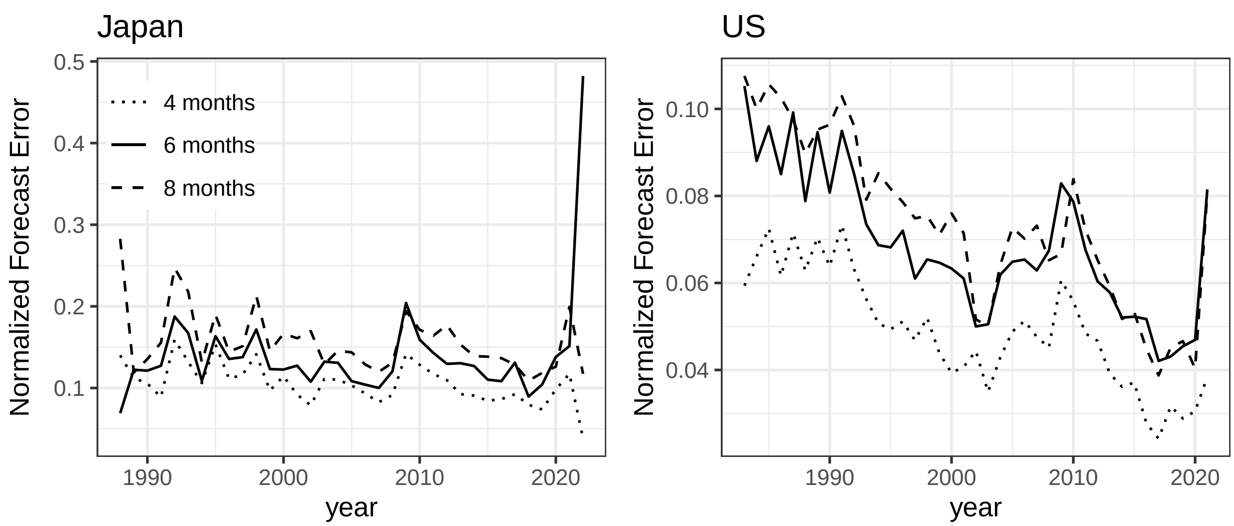


Figure 8: Normalized Forecast Error for Earnings-per-Share Forecasts for the Median Firm Notes: The dotted, solid, and dashed lines represent the normalized errors for forecasts made 8, 6, and 4 months before the fiscal year end date for the median stock, respectively.

U.S. We restrict attention to years where there are at least 40 stocks with monthly price data that allows the estimation of range volatility. Both measures are relatively stable, with high volatility episodes in 1987, 1998, 2001, 2008, and 2020.

The experience of the median firm is representative of the majority of the firms in both stock markets. Table 9 provides the cross-sectional summary statistics in 2018 for Japan and the U.S. The interquartile range is similar to the time series variation for the median firm.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
U.S.	0.01	0.07	0.11	0.13	0.16	2.37
Japan	0.03	0.06	0.11	0.13	0.15	0.74

Table 9: The Summary Statistics for the Normalized Range Volatilities in 2018

C.3 Factor Loading Estimates

In this section, we provide summary statistics for the estimates of the loadings (betas) on the Fama-French factors. Figure 10 depicts the median firm's beta estimates for Japan and the U.S. We restrict attention to years where there are at least 40 stocks with monthly

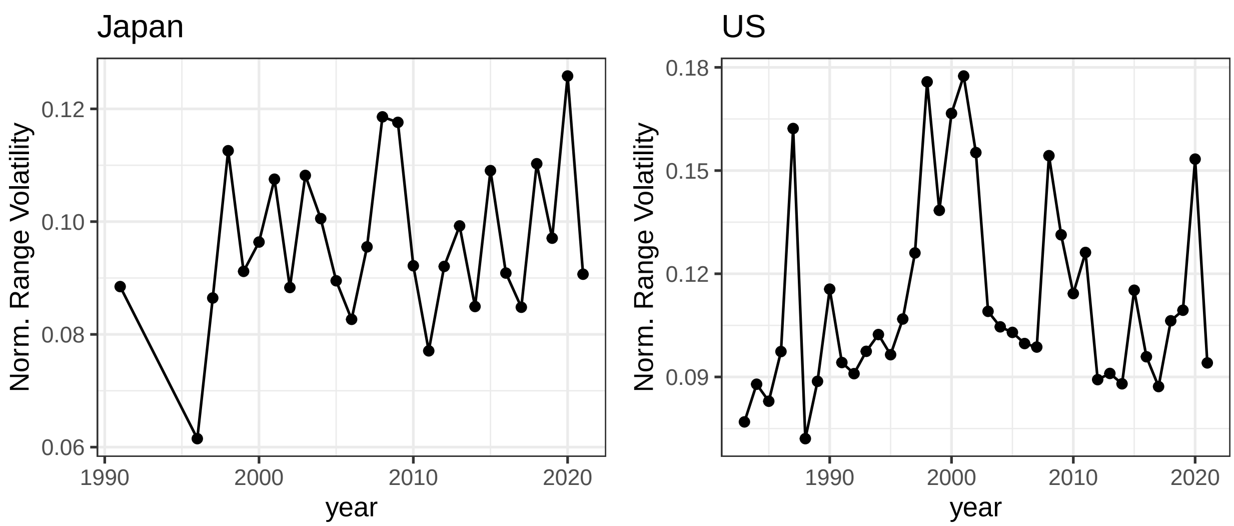


Figure 9: Normalized Range Volatilities for the Median Firm

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
BMarketReturn	-13.32	0.27	0.61	0.71	1.03	12.75
BSMB	-8.19	-0.45	0.15	0.32	0.87	33.66
BHML	-92.53	-0.61	0.06	-0.08	0.68	8.58

Japan Estimates

price data that allows the estimation of range volatility. We can see that the median stock in Japan moves less in tandem with the market compared to the median stock in the U.S. Furthermore, the factor loadings on high-minus-low and small-minus-big factors are more volatile.

Table 10 provides the cross-sectional summary statistics on beta estimates from both countries. The U.S., in particular, has many outliers, which are winsorized before being used in the pricing regressions. In practice, the Fama-French estimation is generally used for portfolios instead of individual stocks. Hence, the cross-sectional dispersion of betas here is larger than the estimates generally found in the literature.

C.4 Distribution of signals and forecast errors: United States

Here, we report the parameters determining the distribution of signals and forecast errors for the United States as an illustration of the results. Figure 11 shows the time series of

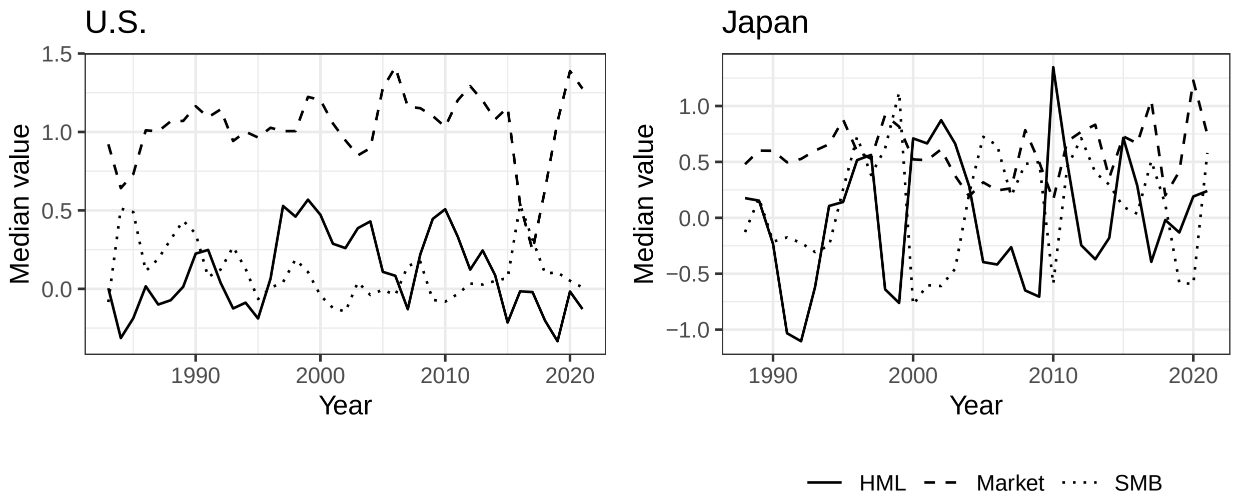


Figure 10: Factor Loadings for the Median Stock from the 3-Factor Fama-French Estimation Notes: The dashed, solid, and dotted lines refer to the beta estimates for the Market return, High-Minus-Low, and Small-Minus-Big factors.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
BMarketReturn	-17637.19	0.55	1.05	0.63	1.61	2373.02
BSMB	-4054.64	-0.22	0.08	0.16	0.47	4137.85
BHML	-7669.93	-0.45	0.04	-0.11	0.73	5453.80

The U.S. Estimates

Table 10: Summary Statistics for Fama-French 3-Factor Loadings in 2018

the forecast error variance ($\sigma^2_{\varepsilon_n}$ and $\sigma^2_{\varepsilon_d}$). Figure 12 shows the time series of the signal averages ($\bar{\theta_n}$ and $\bar{\theta_d}$). Figure and 13 shows ($\sigma^2_{\theta_n}$ and $\sigma^2_{\theta_d}$) that are needed to construct the measure of price informativeness.

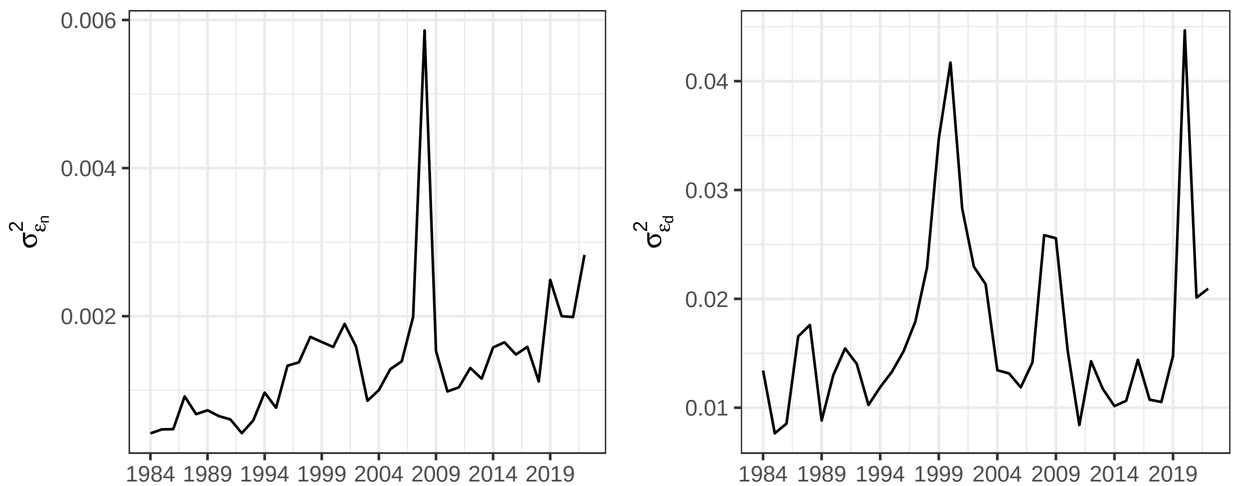


Figure 11: Forecast Error Variance Estimates in the U.S. for a Balanced Panel

Figure 12: Median Earnings and Volatility Signals in the U.S. for a Balanced Panel

D Additional Figures and Tables

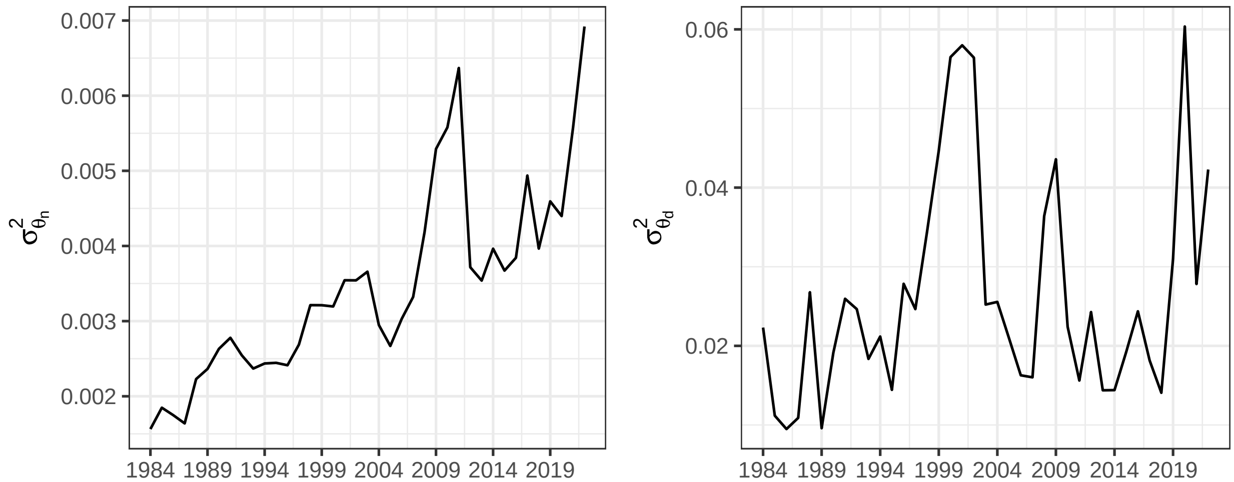


Figure 13: Variances of the Earnings and Volatility Signals in the U.S. for a Balanced Panel

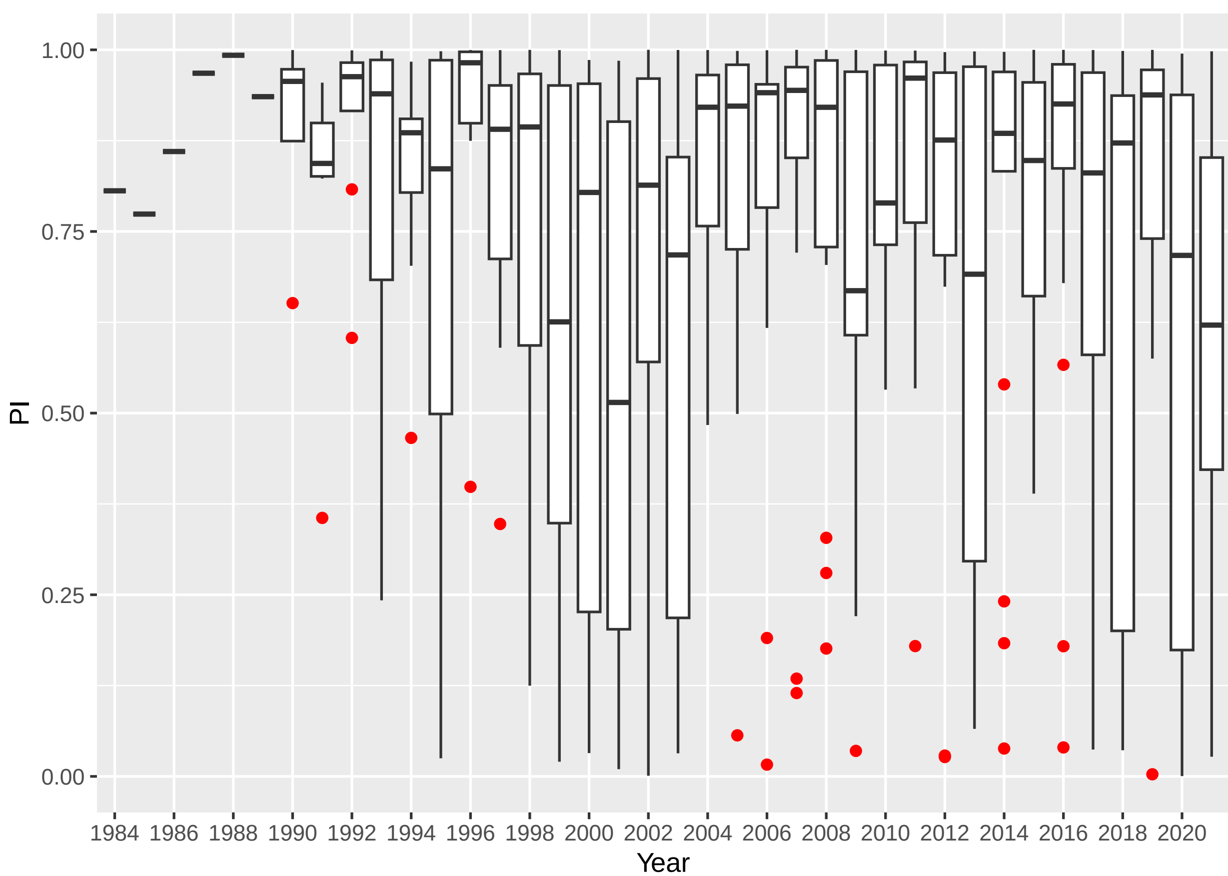


Figure 14: Yearly Box-plots of Price Informativeness Estimates