Identifying Price Informativeness*

Eduardo Dávila[†] Cecilia Parlatore[‡]

June 2019

Abstract

We show that outcomes (parameter estimates and R-squareds) of regressions of prices on fundamentals allow us to recover *exact* measures of the ability of asset prices to aggregate dispersed information. We formally show how to recover absolute and relative price informativeness in dynamic environments with rich heterogeneity across investors (regarding signals, private trading needs, or preferences), minimal distributional assumptions, multiple risky assets, and allowing for stationary and non-stationary asset payoffs. We implement our methodology empirically, finding stock-specific measures of price informativeness for U.S. stocks. We find a right-skewed distribution of price informativeness, measured in the form of the Kalman gain used by an external observer that conditions its posterior belief on the asset price. The recovered mean and median are 0.05 and 0.02 respectively. We find that price informativeness is higher for stocks with higher market capitalization and higher trading volume.

JEL Classification: D82, D83, G14

Keywords: price informativeness, information aggregation, informational efficiency, model identification, Hayek

*We would like to thank Philip Bond, Ricardo Caballero, Jennifer Carpenter, John Campbell, Ian Dew-Becker, Darrell Duffie, Emmanuel Farhi, Itay Goldstein, Gary Gorton, Joel Hasbrouck, Arvind Krishnamurthy, Ben Lester, Stephen Morris, Giuseppe Moscarini, Stefan Nagel, Guillermo Ordoñez, Michael Sockin, David Thesmar, Aleh Tsyvinski, Kathy Yuan, Wei Xiong, Laura Veldkamp, and Toni Whited, as well as seminar participants at SITE, Michigan Ross, BYU, University of Luxembourg, University of Frankfurt, NYU Five Star, UC Berkeley Haas, Yale, University of Virginia, University of Chicago, Sciences Po, Adam Smith Conference, Stockholm School of Economics, University of Washington Foster, and FIRS, for helpful comments and discussions. We are especially thankful to Alexi Savov for extended conversations on the topic of this paper, as well as for sharing code and data. Luke Min provided outstanding research assistance.

[†]Yale University and NBER. Email: eduardo.davila@yale.edu

[‡]New York University, Stern School of Business. Email: cparlato@stern.nyu.edu

1 Introduction

Financial markets play an important role by aggregating dispersed information about the fundamentals of the economy. By pooling different sources of information, asset prices act as a public signal to any external observer, potentially influencing individual decisions. This view, uncontested within economics and traced back to Hayek (1945), has faced significant challenges when translating its theoretical findings to more applied settings because measuring the informational content of prices is not an easy task.[1] In particular, one may be interested in understanding whether different markets aggregate dispersed information to different degrees. However, without direct measures of the informational content of prices (price informativeness), how can one know which markets are better at aggregating information?

In this paper, we develop a methodology that allows us to recover *exact* stock-specific measures of price informativeness. Formally, we show that the outcomes (parameters and R-squareds) of linear regressions of prices on fundamentals are sufficient to identify absolute and relative price informativeness within a large class of linear asset demand models that feature rich heterogeneity across investors (regarding signals, private trading needs, or preferences) and minimal distributional assumptions.[2]

Throughout the paper, we formally define and study two measures of the informational content of prices: absolute and relative price informativeness. Absolute price informativeness measures the amount of information contained in the price for an investor who only learns about an asset's payoff from the price. We formally define absolute price informativeness as the precision of the unbiased signal about the innovation to the fundamental revealed by asset prices. Relative price informativeness measures the informational content of prices relative to the total amount of uncertainty about the asset's payoff. Formally, relative price informativeness corresponds to the ratio between absolute price informativeness and the precision of the innovation to the fundamental, which measures the underlying source of uncertainty. This measure captures how much can be learned from the price relative to the total amount that can be learned.

The main contribution of this paper is methodological. For illustration, let us describe how price informativeness can be recovered. In a stationary environment, consider running the following regression relating the ex-dividend price at which an asset is traded in period t, p_t, to the asset payoff to be realized at the end of period t, θ_{t+1}, and its contemporary payoff, θ_t,

$$p_t = \beta_0 + \beta_1 \theta_t + \beta_2 \theta_{t+1} + \varepsilon_t. \tag{R0}$$

The equilibrium relation that supports Regression R0 provides the foundation for our procedure to identify price informativeness. We show that $\frac{\beta_2^2}{\mathbb{V}\text{ar}[\varepsilon_t]}$ exactly corresponds to absolute price informativeness, that is, to the precision of the unbiased signal about the innovation to the fundamental

[1] Hayek (1945) highlights the relevance of price informativeness as follows: "The economic problem of society is (...) rather a problem of how to secure the best use of resources known to any of the members of society, for ends whose relative importance only these individuals know. Or, to put it briefly, it is a problem of the utilization of knowledge which is not given to anyone in its totality."

[2] Given that linear asset demands can be interpreted as an approximation to more general models, one should expect our results to be valid more broadly in an approximate sense.

contained in asset prices. We structurally map the error term ε_t to the noise component of asset prices, and we show how it can arise from different primitive assumptions (e.g. random heterogeneous priors, multiple risky assets, unlearnable component of the fundamental, etc.).

We also show that the difference between the R-squared of Regression R0 and the R-squared of an identical regression that does not include θ_{t+1} can be directly mapped to relative price informativeness. Formally, we show that a corrected incremental R-squared between both regressions exactly determines relative price informativeness. Under Gaussian primitives, we also show that the same difference directly determines the Kalman gain that an external observer attributes to the innovation to the fundamental when forming a posterior about the future payoff of the asset. Importantly, our results are not only qualitative, but they provide exact measures of informativeness. For instance, finding a Kalman gain of $K = 0.2$ implies that an external observer weighs the information contained in the price by 20% relative to his prior.

It is worth highlighting that our identification procedure does not make parametric assumptions regarding the underlying source of noise and does not require that investors form beliefs using Bayesian updating. We also discuss in detail how our results relate to alternative measures that relate to price informativeness, like the posterior variance of the fundamental or forecasting price efficiency measures.

Although we derive our identification results without the need to fully specify the model primitives, we explicitly develop a fully microfounded dynamic model of trading in Section 4 of the paper. In the context of this model, we provide a new identification result that allows us to recover, using aggregate information, the precision of investors' private signals and the volatility of the aggregate component of investors trading needs (noise). To our knowledge, this result provides the first methodology that transparently recovers the precision of investors' signals precision in REE models. It also provides a direct methodology to capture the amount of noise trading in REE models.

We systematically extend our methodology to more general environments. In particular, we show how to adapt our methodology to allow for unit-roots in the process followed by the fundamental. When allowing for non-stationary payoffs, we show how to implement our results in difference form. We also extend our results to include multiple risky assets, payoffs with learnable and unlearnable components, and public signals about the asset payoff. In all scenarios, our methodology remains valid to answer the question of how much an external observer can learn from the price. These extensions highlight that the exact interpretation of the noise term depends on the exact assumptions of the model. Throughout the paper, we discuss at length the rationale behind our measures of informativeness, and how the exercise of identifying price informativeness is distinct from exploring predictability relations.

After describing the methodology, we proceed to implement our results empirically, recovering actual measures of price informativeness. Exploiting our theoretical framework, we run regressions of prices on fundamentals at the stock level to recover exact measures of absolute and relative price informativeness. Crucially, by exploiting time series variation for a given stock, we are able to recover exact measures of price informativeness that are asset-specific. Our theoretical results allow us to provide an exact structural interpretation of our estimates and their magnitudes.

To implement our identification results empirically, we use stock market values and quarterly earnings as a measure of fundamentals. Using data from 1963 to 2017, compute exact measures of price informativeness for each individual stock in our sample using our identification results. Our empirical implementation generates distributions of absolute price informativeness, relative price informativeness, and Kalman gains across stocks. We find a right skewed distribution of Kalman gains, with a mean and median Kalman gain of 0.05 and 0.02. These numbers imply that, for half of the stocks in the sample, the information contained in the price would weight less than 2% in the posterior of a Bayesian investor. The magnitudes of these estimates suggest that prices contain substantially more noise than information.

The distribution of stock-specific estimates exhibits substantial cross-sectional dispersion. This heterogeneity in price informativeness at the stock level is ubiquitous across stock characteristics such as the exchange in which they are traded, their market capitalization, the volume traded and their corresponding industry. We find that price informativeness is higher for stocks traded in the NYSE, with higher market capitalization, and traded more frequently.

Related Literature Our theoretical framework builds on the literature that studies the role played by financial markets in aggregating dispersed information, following Grossman and Stiglitz (1980), Hellwig (1980), and Diamond and Verrecchia (1981), De Long et al. (1990), among others. Vives (2008) and Veldkamp (2009) provide thorough reviews of this well-developed and growing body of work.

While the role of financial markets aggregating information has been the subject of a substantial theoretical literature, the development of empirical measures of price informativeness is more recent. The work of Bai, Philippon and Savov (2015) is perhaps the closest. Leaving aside that we consider a substantially richer framework than theirs, there are three significant differences between their approach and ours. First, we focus on the ability of financial markets to aggregate information while their focus is on the allocation of capital. As we show in the paper, the measure they use to make inferences about price informativeness (forecasting price efficiency, V_{FPE}) does not allow to separately identify the role of financial markets aggregating dispersed information from the volatility of the fundamental. While their measure V_{FPE} is relevant to infer whether the allocation of capital in financial markets has improved or worsened, our results show that it is not the right measure to understand whether financial markets have become better at aggregating information. Forecasting price efficiency, V_{FPE}, can be high due to a low volatility of the fundamental or due to a high level of price informativeness. Second, since their focus is on the forecasting power of prices (about future fundamentals), they focus on a regressions of fundamentals on asset prices. Instead, we focus on the relation between an endogenous variable (price) and exogenous variables (fundamentals), which avoids potential biases in the estimation. Finally, they estimate V_{FPE} by running cross-sectional regressions, which implicitly assumes that the fundamental and noise are distributed identically across all of the stocks in their sample. Instead, we exploit time series variation to recover asset specific measures of informativeness.

Following the approach in Bai, Philippon and Savov (2015), Farboodi, Matray and Veldkamp (2017) also use forecasting price efficiency to infer changes in price informativeness. They find that while average price informativeness increased in the S&P500, it decreased for the whole sample. They argue that this can be due to a composition effect in the S&P500 as price informativeness increased for large firms and decreased for small ones. Kacperczyk, Sundaresan and Wang (2018) find that forecasting price efficiency increases with ownership by foreign investors.

There exists earlier work that proposes ad-hoc variables to study the informational content of prices. Influenced by the predictions of the CAPM/APT frameworks and following Roll (1988), Morck, Yeung and Yu (2000) study regressions of returns on a single or multiple factors and informally argue that the R^2 of such regressions can be used to capture whether asset prices are informative/predictive about firm-specific fundamentals. This measure, sometimes referred to as price nonsynchronicity, has been used in several empirical studies that link price informativeness to capital allocation. Wurgler (2000) finds that countries with higher price nonsynchronicity display a better allocation of capital. Durnev, Morck and Yeung (2004) document a positive correlation between price nonsynchronicity and corporate investment. Chen, Goldstein and Jiang (2006) show a positive relation between the sensitivity of corporate investment to price and two measures of the information contained in prices, price nonsynchronicity and the probability of informed trading (PIN), and conclude that managers learn from the price when making corporate investment decisions.[3]

More recently, Weller (2018) uses a price jump ratio to measure how much information enters prices relatively to how much is potentially acquirable at the stock level. Using this measure, he finds that algorithmic trading decreases the amount of information that is incorporated in prices.

While these results uncover interesting empirical relations, the measures used by this body of work do not have a structural interpretation. Hou, Peng and Xiong (2013) forcefully highlight the importance of this structural link in the context of the return R^2. They question the link between return R^2 and price informativeness theoretically, in rational and behavioral settings, and empirically. Moreover, even if the existing measures may correlate with price informativeness, it is impossible to translate the magnitude of the changes in these variables into changes in the informational content of prices without a structural interpretation. By showing how to recover exact measures of stock specific price informativeness, we can reach quantitative conclusions about price informativeness.

As in any structural model, the measure of informativeness that we recover is linked to our assumptions on the behavior of investors and the market structure. While our framework is general along several dimensions, there is scope to think about how to identify price informativeness in alternative models of trading that depart from our linearity assumption. In particular, our analysis purposefully abstracts from feedback between prices and fundamentals, summarized in Bond, Edmans and Goldstein (2012), and tested in and Chen, Goldstein and Jiang (2006), which introduces fundamental non-linearities that can only be considered using full-information methods.

Outline Section 2 describes the assumptions of our baseline model, formally introduces the definitions of absolute and relative price informativeness, and presents our main results regarding how to recover

[3]The PIN, developed in Easley, O'Hara and Paperman (1998), seeks to measures the probability of an informed trade in a model with noise and informed traders.

price informativeness from linear regressions. Section 3 introduces a fully microfounded model that maps to the assumptions on endogenous objects made in Section 2 and that allows us to fully recover model primitives. Section 4 extends our methodology to more general environments, including non-stationary payoffs, multiple risky assets, and payoffs with learnable and unlearnable components. Section 5 empirically implements the methodology introduced in the paper, while Section 6 concludes. All proofs and derivations are in the Appendix.

2 Identifying Price Informativeness

In this section, we introduce the main identification results in the context of a dynamic model with a single risky asset whose payoff process is stationary. We extend the results to more general environments in Section 4.

2.1 Model

Time is discrete, with periods denoted by $t = 0, 1, 2, \ldots, \infty$. There is a continuum of investors, indexed by $i \in I$, who trade a risky asset in fixed supply each period at a price p_t. The payoff of the risky asset in period $t+1$, θ_{t+1}, is given by the following stationary AR(1) process

$$\theta_{t+1} = \mu_\theta + \rho\theta_t + \eta_t,$$

where μ_θ is a scalar, $|\rho| < 1$, and where the innovations to the payoff, η_t, have mean zero, finite variance, and are independently distributed. Investors trade in period t with imperfect information about the innovation to the payoff, η_t, which is realized at the end of the period. When trading in period t, the contemporaneous payoff θ_t has already been realized and is common knowledge to all investors. We often refer to the asset payoff θ_{t+1} as the fundamental.

Each period t, an investor i observes a private signal s_t^i of the innovation to the payoff η_t.[4] Investors have an additional motive for trading the risky asset that is orthogonal to the asset payoff. We denote by n_t^i investor i's additional trading motive in period t. These additional trading motives are private information of each investor and are potentially random in the aggregate.

We derive the main results of the paper under two assumptions. The first assumption imposes an additive informational structure and guarantees the existence of second moments, while the second assumption imposes a linear structure for investors' equilibrium asset demands. In general, linear demands can be interpreted as a first-order approximation to other forms of asset demands, so one may expect our results to approximately hold in a larger class of models. Both assumptions facilitate the aggregation of individual demands in order to yield a linear equilibrium pricing function.

Assumption 1. (Additive noise) *Each period t, every investor i receives an unbiased private signal s_t^i about the innovation to the payoff, η_t, of the form*

$$s_t^i = \eta_t + \varepsilon_{st}^i, \tag{1}$$

[4] Assuming that investors observe private signals about the payoff, θ_{t+1}, or its innovation, η_t, is formally equivalent, since θ_t is known to investors when trading in period t.

where ε_{st}^i, $\forall i \in I$, $\forall t$ *, are random variables with mean zero and finite variances, whose realizations are independent across investors and over time. Each period* t*, every investor* i *has a private trading need* n_t^i*, of the form*

$$n_t^i = n_t + \varepsilon_{nt}^i, \tag{2}$$

where n_t *is a random variable with finite mean, denoted by* μ_n*, and finite variance, and where* ε_{nt}^i, $\forall i \in I$, $\forall t$ *, are random variables with mean zero and finite variances, whose realizations are independent across investors and over time.*

Assumption 1 imposes restrictions on the noise structure in the signals about the innovation to the fundamental η_t and on all other sources of investors' private trading needs by making them additive and independent across investors. This assumption does not restrict the distribution of any random variable beyond the existence of finite first and second moments. Our second assumption describes the structure of the investors' net demands for the risky asset Δq_t^i.

Assumption 2. (Linear asset demands) *Investors' net asset demands satisfy*

$$\Delta q_t^i = \alpha_s^i s_t^i + \alpha_\theta^i \theta_t + \alpha_n^i n_t^i - \alpha_p^i p_t + \psi^i,$$

where α_s^i, α_θ^i, α_n^i, α_p^i, *and* ψ^i *are individual demand coefficients, determined in equilibrium.*

Assumption 2 imposes a linear structure on the individual investors' net asset demand for the risky asset. More specifically, that an individual investor's net demand is linear in his signal about the fundamental and his private trading needs, as well as in the asset price p_t and the current realization of the fundamental θ_t. It also allows for an individual specific invariant component ψ^i. In Section 3, we provide a fully specified model that is consistent with Assumptions 1 and 2 and briefly describe other models that are consistent with both assumptions.

Given our assumptions, market clearing for the risky asset implies that $\int \Delta q_t^i di = 0$, $\forall t$. Market clearing, exploiting a Law of Large Numbers, yields an equilibrium pricing equation of the form

$$p_t = \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\eta_t + \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t + \frac{\overline{\alpha_n}}{\overline{\alpha_p}}n_t + \frac{\overline{\psi}}{\overline{\alpha_p}}, \tag{3}$$

where we denote the cross sectional averages of individual demand coefficients by $\overline{\alpha_s} = \int_I \alpha_s^i di$, $\overline{\alpha_p} = \int_I \alpha_p^i di$, $\overline{\alpha_\theta} = \int_I \alpha_\theta^i di$, $\overline{\alpha_n} = \int_I \alpha_n^i di$, and $\overline{\psi} = \int_I \psi^i di$. The linearity of net demands implies that the equilibrium asset price is also linear in the future realization of the fundamental θ_{t+1}, the current fundamental θ_t, and the common component of investors' private trading needs n_t. Without loss of generality, we assume that n_t and η_t are independent.[5] An interpretation of Equation (3) as a linear regression provides the foundation for our procedure to identify price informativeness.

Although the equilibrium price p_t is the main endogenous observable variable generated by the model, the relevant variable from the perspective of information aggregation is $\hat{p}_t$, defined by

$$\hat{p}_t \equiv \frac{\overline{\alpha_p}}{\overline{\alpha_s}}p_t - \frac{\overline{\alpha_\theta}}{\overline{\alpha_s}}\theta_t - \frac{\overline{\alpha_n}}{\overline{\alpha_s}}\mu_n - \frac{\overline{\psi}}{\overline{\alpha_s}}, \tag{4}$$

[5] In the Appendix, we consider the case in which the aggregate noise n_t is correlated with the fundamental η_t and show that all our identification results remain valid.

which corresponds to the unbiased signal of the innovation η_t contained in the price p_t. Note that $\hat{p}_t$ corresponds to $\hat{p}_t = \eta_t + \frac{\overline{\alpha_n}}{\overline{\alpha_s}}(n_t - \mu_n)$, implying that $\mathbb{E}[\hat{p}_t|\theta_{t+1}, \theta_t] = \eta_t$. Because the contemporary realization of the fundamental θ_t is observed at date t, information about the innovation η_t translates one-for-one to information about the asset payoff. Using the definition of $\hat{p}_t$, we can formally define the two measures of price informativeness that we show how to recover in this paper as follows.[6]

Definition 1. (Absolute price informativeness) *We define absolute price informativeness as the precision of the unbiased signal about the innovation to the fundamental payoff θ_{t+1} contained in the asset price p_t. We denote absolute price informativeness by $\tau_{\hat{p}}$, which formally corresponds to*

$$\tau_{\hat{p}} \equiv (\mathbb{V}ar[\hat{p}_t|\theta_{t+1}, \theta_t])^{-1} = \left(\frac{\overline{\alpha_s}}{\overline{\alpha_n}}\right)^2 \tau_n, \tag{5}$$

where $\tau_n \equiv \mathbb{V}ar[n_t]^{-1}$.

Absolute price informativeness increases when investors trade more aggressively on their private signals (high $\overline{\alpha_s}$), when they trade less aggressively on their private trading motives (low $\overline{\alpha_n}$), and when the aggregate component of trading motives is less volatile (has a high precision τ_n). Absolute price informativeness reveals, for given realizations of the future and current fundamentals θ_{t+1} and θ_t, the possible dispersion of observed equilibrium prices. In a statistical sense, it indicates whether the signal contained in the price is close to the fundamental. Consequently, absolute price informativeness captures how much information about the fundamental can be gained by an uninformed external observer by exclusively observing the price. When absolute price informativeness is high, an external observer receives a very precise signal about the fundamental by observing the asset price p_t. On the contrary, when price informativeness is low, an external observer learns little about the fundamental by observing the asset price p_t.

Definition 2. (Relative price informativeness) *We define relative price informativeness as the ratio between absolute price informativeness and the precision of the innovation to the fundamental. We denote relative price informativeness by $\tau^R_{\hat{p}}$, which formally corresponds to*

$$\tau^R_{\hat{p}} \equiv \frac{\tau_{\hat{p}}}{\tau_\eta}, \tag{6}$$

where $\tau_\eta \equiv \mathbb{V}ar[\eta_t]^{-1}$.

Relative price informativeness simply corrects absolute price informativeness for the precision of the innovation to the fundamental. This measure expresses how much can be learned by observing the price relative to the volatility of the fundamental. As we will show below, there is a tight connection between relative price informativeness and the Kalman gain of an external observer in Gaussian models with Bayesian updating.

We would like to conclude the description of the model environment with the following remarks.

[6] The measures of price informativeness that we study in this paper are the relevant measures for an external observer who learns about the fundamental from the price. See Davila and Parlatore (2018) for a discussion on how to link external price informativeness to internal price informativeness, which may be the relevant object of interest for investors within the model in some environments. That paper systematically studies the subtle relation between the volatility of the equilibrium price p_t and the precision of the unbiased signal about the fundamental innovation $\hat{p}_t$.

Remark 1. *Precision of price as signal of fundamentals versus posterior variance.* The variance of the signal about the innovation to the fundamental has the advantage that it can be derived without the need to make assumptions about how an external observer updates its information about the fundamental. In our setup, it is possible to calculate $\mathbb{V}\text{ar}\left[\hat{p}_t|\theta_{t+1}, \theta_t\right]$ without making distributional assumptions beyond the existence of second moments. However, to calculate the posterior variance $\mathbb{V}\text{ar}\left[\theta_{t+1}|\hat{p}_t, \theta_t\right]$, it is necessary to make assumptions regarding the distribution of priors and signals. For this reason, $\tau_{\hat{p}}$ as defined in Eq. (5) is a more appealing measure of informativeness, since it can be derived (and, as we show in this paper, recovered from observables) without specifying the nature of updating/filtering used by investors. We further discuss the adequacy of the measures just defined and other measures of informativeness after introducing our main results in Section 3.

Remark 2. *Cross-sectional Heterogeneity.* Our framework allows for a rich cross-sectional heterogeneity among investors. In particular, it accommodates heterogeneity in investors' risk aversion, in the precision of their information, and in the distribution of their idiosyncratic trading motives. For instance, our assumptions can accommodate models with informed and uninformed traders, which can be mapped to environments in which one set of agents does not observe any private signal, and those with classic noise traders, which can be mapped to environments in which one set of agents trades fixed amounts regardless of the price or other features of the environment.

Remark 3. *Distributional Assumptions.* It's worth highlighting that Assumption 1 does not require normality of signals or fundamentals, so our main results in Propositions 1 and 2 do not rely on distributional assumptions beyond the existence of well-defined first and second moments. However, at times, we discuss how our results can be more easily interpreted if we assume that signals and fundamentals have a Gaussian structure – we explicitly state in the text for which results/interpretations normality is needed.

Remark 4. *Multiple Assets/General Shock Structure.* For clarity, we introduce our results in the context of a single asset model. We show in Section 4.2 how to reinterpret our results when investors can trade many risky assets with payoff processes that are potentially correlated across assets and with the aggregate trading needs. The more general framework studied in Section 4.2 shows how to reinterpret our results when there are multiple risky assets. In Section 4.1, we extend the methodology to non-stationary processes. We also show in Sections 4.3 how to interpret our results when the payoff features a learnable and an unlearnable component.

2.2 Identification Results

Exploiting Assumptions 1 and 2, as well as market clearing, we now proceed to derive Propositions 1 and 2, which provide the main identification results of the paper. We can rewrite the equilibrium price introduced in Equation (3) in terms of the future and contemporary asset payoffs as follows:

$$p_t = \left(\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}} - \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\rho\right)\theta_t + \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\theta_{t+1} + \frac{\overline{\alpha_n}}{\overline{\alpha_p}}n_t + \frac{\overline{\psi}}{\overline{\alpha_p}}. \tag{7}$$

This reformulation of the equilibrium pricing equation allows us to identify price informativeness from measures of prices and fundamentals. We sequentially show how to use the outcomes (coefficients and R-squareds) of a regression of prices on fundamentals to exactly recover measures of absolute and relative price informativeness.

Proposition 1. (Identifying absolute price informativeness) *Assume that the additive noise assumption and the linear asset demands assumption are satisfied. Let β_0, β_1, and β_2 denote the coefficients of the following regression of prices on fundamentals,*

$$p_t = \beta_0 + \beta_1 \theta_t + \beta_2 \theta_{t+1} + \varepsilon_t, \tag{R1}$$

where p_t denotes the ex-dividend price at the beginning of period t, θ_{t+1} denotes the measure of fundamentals realized over period t, and where we denote the variance of the error by $\sigma_\varepsilon^2 = \mathbb{V}ar[\varepsilon_t]$. Then, absolute price informativeness, $\tau_{\hat{p}}$, can be recovered by

$$\tau_{\hat{p}} = \frac{\beta_2^2}{\sigma_\varepsilon^2}. \tag{8}$$

The proof of Proposition 1 relies on finding the right combination of parameters in our econometric specification defined in Regression R1, that maps into the definition of absolute price informativeness, $\tau_{\hat{p}}$. By comparing Equation (7) with Regression R1, it is easy to verify that

$$\frac{\beta_2^2}{\sigma_\varepsilon^2} = \frac{\left(\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\right)^2}{\left(\frac{\overline{\alpha_n}}{\overline{\alpha_p}}\right)^2 \tau_n^{-1}} = \left(\frac{\overline{\alpha_s}}{\overline{\alpha_n}}\right)^2 \tau_n = \tau_{\hat{p}},$$

which proves our statement. Intuitively, a strong co-movement between the price p_t and the fundamental θ_{t+1} (high β_2) and a high explanatory power of the regression (low σ_ε^2) indicates that prices are more informative. Since the error in Regression R1 is orthogonal to the regressors, OLS provides consistent estimates of β_2 and σ_ε^2, and consequently of $\tau_{\hat{p}}$.[7]

It may seem that Proposition 1 allows us to recover relative price informativeness by computing and dividing by the precision of the innovation to the fundamental τ_η. However, this procedure would require to directly estimate the process for θ_{t+1}, which would require parametric assumptions on the process. Instead, in Proposition 2, we show how to directly recover relative price informativeness, exclusively as a function R-squareds of regressions of prices on fundamentals.

[7]Formally, when $\tilde{\beta}_2$ and $\tilde{\sigma}_\varepsilon^2$ denote consistent estimates of β_2 and σ_ε^2 in Regression R1, $\tau_{\hat{p}}$ can be consistently estimated as $\tilde{\tau}_{\hat{p}} = \frac{\tilde{\beta}_2^2}{\tilde{\sigma}_\varepsilon^2}$, since

$$\operatorname{plim}\left(\tilde{\tau}_{\hat{p}}\right) = \operatorname{plim}\left(\frac{\tilde{\beta}_2^2}{\tilde{\sigma}_\varepsilon^2}\right) = \frac{\beta_2^2}{\sigma_\varepsilon^2} = \frac{\left(\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\right)^2}{\left(\frac{\overline{\alpha_n}}{\overline{\alpha_p}}\right)^2 \tau_n^{-1}} = \left(\frac{\overline{\alpha_s}}{\overline{\alpha_n}}\right)^2 \tau_n = \tau_{\hat{p}}.$$

Throughout the paper we distinguish between economic identification, understood as the ability to recover model parameters or other endogenous objects of interest from observable variables, and consistent estimation. The focus of this paper is on identification, although at times with discuss the consistency properties of standard estimators when implementing the results of our model.

Proposition 2. (Identifying relative price informativeness) *Assume that the additive noise and the linear asset demands assumptions are satisfied. Let* $R^2_{|\theta_{t+1},\theta_t} \equiv 1 - \frac{\mathbb{V}ar[\varepsilon_t]}{\mathbb{V}ar[p_t]}$ *be the R-squared of Regression R1. Let* $R^2_{|\theta_t}$*,* ζ_0*, and* ζ_1 *respectively denote the R-squared and the coefficients of the following regression of prices on lagged fundamentals*

$$p_t = \zeta_0 + \zeta_1 \theta_t + \varepsilon_t^{\zeta}. \tag{R2}$$

Then, relative price informativeness $\tau^R_{\hat{p}}$ *can be recovered by*

$$\tau^R_{\hat{p}} = \frac{\tau_{\hat{p}}}{\tau_\eta} = \frac{R^2_{|\theta_{t+1},\theta_t} - R^2_{|\theta_t}}{1 - R^2_{|\theta_{t+1},\theta_t}}.$$

Note that the recovered value of relative price informativeness $\tau^R_{\hat{p}}$ has to be non-negative, since $R^2_{|\theta_{t+1},\theta_t} \geq R^2_{|\theta_t}$ and $R^2_{|\theta_{t+1},\theta_t} \in [0,1]$. Intuitively, $\tau^R_{\hat{p}}$ is increasing on $R^2_{|\theta_{t+1},\theta_t}$ for two reasons: a higher $R^2_{|\theta_{t+1},\theta_t}$ reflects a lower residual uncertainty after observing the price and accounting for the lagged fundamental ($\left(1 - R^2_{|\theta_{t+1},\theta_t}\right)$ is lower) and a larger reduction in uncertainty after observing the price relative to only accounting for the lagged fundamental ($\left(R^2_{|\theta_{t+1},\theta_t} - R^2_{|\theta_t}\right)$ is higher). By directly relying on Proposition 2 there is no need to directly estimate τ_η or the AR coefficient ρ.

It may be helpful to relate relative price informativeness to the Kalman gain used by an outside observer. In particular, if we further assume that all primitive random variables are Gaussian, a non-linear transformation of relative price informativeness maps to the Kalman gain of an external Bayesian observer, as expressed in the following corollary to Proposition 2.[8]

Corollary. (Kalman gain) *Under a Gaussian information structure, the Kalman gain for an external Bayesian observer, denoted by* K*, can be recovered as follows*

$$K \equiv \frac{\tau_{\hat{p}}}{\tau_{\hat{p}} + \tau_\eta} = \frac{\tau^R_{\hat{p}}}{1 + \tau^R_{\hat{p}}} = \frac{R^2_{|\theta_{t+1},\theta_t} - R^2_{|\theta_t}}{1 - R^2_{|\theta_t}}.$$

Note that Kalman gains must take values between 0 and 1. This feature makes them appealing from the perspective of interpreting the results. When $R^2_{|\theta_{t+1},\theta_t} \to 1$, the posterior of an external observer fully disregards the prior information, putting all the weight on the price as a signal of the innovation. On the contrary, when $R^2_{|\theta_{t+1},\theta_t} \to R^2_{|\theta_t}$, an external observer does not update his prior at all after observing the price. In the special case in which the payoff process is i.i.d., so the lagged fundamental is irrelevant to predict future fundamentals, $R^2_{|\theta_t} \to 0$, and the R-squared of a regression of prices on fundamentals exactly maps to the Kalman gain of an external observer.

Figure 1 illustrates how to graphically interpret the recovered Kalman gain. Intuitively, the denominator $1 - R^2_{|\theta_t}$ can be interpreted as the share of uncertainty to be learned after accounting

[8] The Kalman gain corresponds to the relative weight given to the price as a signal about the fundamental. Formally, the posterior distribution of an external observer who makes use of the price as a signal about the innovation to the fundamental is given by

$$\eta_t | p_t \sim N\left(K \cdot \hat{p}_t, (\tau_{\hat{p}} + \tau_\eta)^{-1}\right), \quad \text{where} \quad K = \frac{\tau_{\hat{p}}}{\tau_{\hat{p}} + \tau_\eta},$$

and $\hat{p}_t$ corresponds to Equation (4).

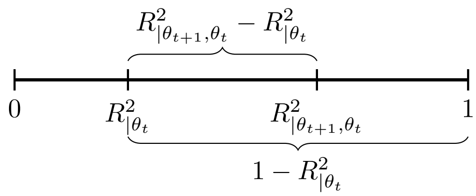

Figure 1: Kalman gain interpretation

for the contemporary fundamental. The numerator can be interpreted as the share of information learned by conditioning on the price relative to the contemporary fundamental. The Kalman gain corresponds to the fraction of an external observer's precision about the fundamental that is conveyed by observing the price. For instance, a Kalman gain of 0.4 implies that 40% of investors' ex-post precision about the innovation to the fundamental comes from conditioning on the price.

Together, Propositions 1 and 2 show that the outcomes of regressions of prices on fundamentals are sufficient to directly recover exact measure of absolute and relative price informativeness in environments with rich heterogeneity across investors (regarding signals, private trading needs, or preferences) and minimal distributional assumptions.

In the remainder of this section, we would like to make two remarks concerning our main results.

Remark 5. *(Alternative measures of informativeness)* Although we have advocated for the precision of the unbiased signal about the fundamental as an appropriate measure to assess the role of prices aggregating information, one can consider other measures. Two alternative measures are i) the *posterior variance* of the fundamental given the price, that is, $\mathcal{V}_P \equiv \mathbb{V}\mathrm{ar}\left[\theta_{t+1}|\theta_t, p_t\right]$, and ii) *forecasting price efficiency* (FPE), which is given by $\mathcal{V}_{FPE} \equiv \mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\theta_{t+1}|\theta_t, p_t\right]\right]$.[9] Under Gaussian uncertainty and Bayesian updating, both measures can be expressed as

$$\mathcal{V}_P = (\tau_\eta + \tau_{\hat{p}})^{-1} = \frac{1}{1+\tau^R_{\hat{p}}}\frac{1}{\tau_\eta} \qquad \text{and} \qquad \mathcal{V}_{FPE} = \frac{\tau^R_{\hat{p}}}{1+\tau^R_{\hat{p}}}\frac{1}{\tau_\eta}. \tag{9}$$

From Equation (9), it is evident that both measures face the same challenge: They confound the effect of uncertainty about the fundamental with price informativeness. For instance, $\mathcal{V}_{FPE}$, which is the measure of informativeness used by Bai, Philippon and Savov (2015) and subsequent work, can be be high because the fundamental is easy to predict (high τ_η) or because the price is a very precise signal of the fundamental (high $\tau^R_{\hat{p}}$). The same ambiguous inference applies to $\mathcal{V}_P$, suggesting that neither of these measures is adequate to recover price informativeness.

[9] Note that the posterior variance of the fundamental and forecasting price efficiency are two sides of the same coin. While the former measures the residual uncertainty about the fundamental after observing the price, the latter measures how much uncertainty about the fundamental is dissipated by observing the realization of the price. Both measures are linked through the Law of Total Variance, as follows

$$\mathbb{V}\mathrm{ar}\left[\theta_{t+1}|\theta_t\right] = \underbrace{\mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\theta_{t+1}|\theta_t, p_t\right]\right]}_{\mathcal{V}_{FPE}} + \mathbb{E}\left[\underbrace{\mathbb{V}\mathrm{ar}\left[\theta_{t+1}|\theta_t, p_t\right]}_{\mathcal{V}_P}\right].$$

In Bai, Philippon and Savov (2015) $\rho = 0$ and, hence, forecasting price efficiency is defined as $\mathcal{V}_{FPE} \equiv \mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\theta_{t+1}|p_t\right]\right]$.

Remark 6. *(Predictability versus informativeness/Reverse regression)* It's worth highlighting that our goal is not to predict future fundamentals from current prices. Instead, our goal is to understand how good are financial markets at aggregating information. For this reason, it is natural to consider regressions of prices (endogenous variable) on future fundamentals (exogenous), even though this entails considering the regression of a variable observable in period t on a explanatory variable realized in the future. Even if one is not interested in predictability, one may wonder why not run regressions of fundamentals on prices, since this type of regression can be use for predictive purposes. This would imply reinterpreting Regression R1 as follows

$$\theta_{t+1} = \gamma_0 + \gamma_1 \theta_t + \gamma_2 p_t + \nu_t, \tag{R3}$$

where $\gamma_0 = -\frac{\overline{\psi}}{\overline{\alpha_s}}$, $\gamma_1 = \frac{\overline{\alpha_\theta}}{\overline{\alpha_s}} - \rho$, $\gamma_2 = \frac{\overline{\alpha_p}}{\overline{\alpha_s}}$, and $\nu_t = -\frac{\overline{\alpha_n}}{\overline{\alpha_s}} n_t$. The main pitfall of this regression is that OLS estimates of the coefficients and the residual variance will be biased, since $\mathbb{C}\text{ov}\left(p_t, \nu_t\right) = \frac{\overline{\alpha_n}}{\overline{\alpha_s}} \frac{\overline{\alpha_n}}{\overline{\alpha_p}} \mathbb{V}\text{ar}\left(n_t\right) \neq 0$.[10] For this reason, given the question addressed in this paper and the exclusion restrictions imposed by our model, it is more natural to consider regressions similar to R1.

Remark 7. *(Irrelevance of Bayesian updating)* Finally, it's worth noting that the measures of absolute and relative price informativeness do not require that investors update their beliefs by Bayesian updating. This is an important consideration, since it allows for rich patterns of belief formation. See Barberis (2018) and Gennaioli and Shleifer (2018) for recent accounts of the importance of non-fully rational expectation formation. That said, any results presented in the form of Kalman gains rely on the assumption that the external observer faces Gaussian uncertainty and uses Bayesian updating.

3 Fully Specified Environment

In our analysis so far, we have remained agnostic about the source of the noise that is impounded in the price and the way in which investors learn from the price. In this section, we study a particular dynamic learning model with overlapping generations that endogenously satisfies Assumptions 1 and 2. Our goal in describing this particular model is two-fold. First, it provides a tractable framework that maps to the main assumptions on equilibrium objects made in Section 2. Second, it allows us to provide a new identification result. Given the new set of assumptions, we are able to exactly recover, using aggregate information, the precision of investors' private signals and the volatility of the aggregate component of investors trading needs (noise).

3.1 Environment

Time is discrete, with periods denoted by $t = 0, 1, 2, \ldots, \infty$. Each period t, there is a continuum of investors, indexed by $i \in I$. Each generation lives two periods and has exponential utility over their

[10] When $\rho = 0$, the bias of the OLS estimate of γ_1 can be easily calculated: $\tilde{\gamma}_1 = \kappa \frac{\overline{\alpha_p}}{\overline{\alpha_s}}$, where $\kappa = \frac{\left(\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\right)^2 \mathbb{V}\text{ar}(\theta_t)}{\left(\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\right)^2 \mathbb{V}\text{ar}(\theta_t) + \left(\frac{\overline{\alpha_n}}{\overline{\alpha_p}}\right)^2 \mathbb{V}\text{ar}(n_t)}$.

last period wealth. An investor born at time t has preferences given by

$$U\left(w_{t+1}\right)=-e^{-\gamma w_{t+1}},$$

where γ is the coefficient of absolute risk aversion and w_{t+1} is the investor's wealth in his final period. There are two long-term assets in the economy: A risk-free asset in perfectly elastic supply, with return $R>1$, and a risky asset in fixed supply Q. The payoff of the risky asset each period t is given by

$$\theta_{t+1}=\mu_\theta+\rho\theta_t+\eta_t,$$

where μ_θ is a scalar, $|\rho|<1$, and $\theta_0=0$. The dividend θ_t is realized and becomes common knowledge at the end of period $t-1$. The innovation in the dividend, η_t, and, hence, θ_{t+1} are realized and observed at the end on period t. The innovations in the dividend are independently distributed over time.

We assume that investors' private trading needs arise from random heterogeneous priors. This is a particularly tractable formulation that sidesteps many of the issues associated with classic noise trading and that prevents full revelation of information – see Davila and Parlatore (2017) for a thorough analysis of this formulation. Formally, each investor i in generation t has a prior over the innovation at time t given by

$$\eta_t \sim_i N\left(\overline{\eta}_t^i,\tau_\eta^{-1}\right),$$

where

$$\overline{\eta}_t^i=n_t+\varepsilon_{\overline{\eta}t}^i \quad \text{with} \quad \varepsilon_{\overline{\eta}t}^i \overset{\text{iid}}{\sim} N\left(0,\tau_{\overline{\eta}}^{-1}\right)$$

and $n_t\sim N\left(\mu_n,\tau_n^{-1}\right)$ can be interpreted as the aggregate sentiment in the economy, where $n_t\perp\varepsilon_{\overline{\eta}t}^i$ for all t and all i. The aggregate sentiment n_t is not observed and acts as a source of aggregate noise in the economy, preventing the price from being fully revealing.

Each investor i in generation t receives a signal about the innovation in the asset payoff η_t given by

$$s_t^i=\eta_t+\varepsilon_{st}^i \quad \text{with} \quad \varepsilon_{st}^i\sim N\left(0,\tau_s^{-1}\right)$$

and $\varepsilon_{it}\perp\varepsilon_{jt}$ for all $i\perp j$, and $\eta_t\perp\varepsilon_{it}$ for all t and all i.

We focus on stationary equilibria in which demand functions are linear in the price and information set of the investors.

Definition. (Equilibrium) A stationary rational expectations equilibrium in linear strategies is a set of linear demand functions q_t^i for each investor i in generation t and a price function p_t such that: a) q_t^i maximizes investor i's expected utility given his information set for every possible price and b) the price function p_t is such that the market for the risky asset clears each period t, that is, $\int q_t^i di=0$.

There always exists a unique stationary rational expectations equilibrium in linear strategies. In the Appendix, we characterize the equilibrium demand coefficients which are the basis to our procedure to identify the parameters of the model.

3.2 Equilibrium Characterization and Identification Result

As we show in the Appendix, the asset demand submitted by investor i born in period t is given by the solution to the following problem

$$\max_{q_t^i} \left(\mathbb{E}\left[\theta_{t+1} + R^{-1}p_{t+1}|\mathcal{I}_t^i\right] - p_t\right) q_t^i - \frac{\gamma}{2}\mathbb{V}ar\left[\theta_{t+1} + R^{-1}p_{t+1}|\mathcal{I}_t^t\right]\left(q_t^i\right)^2,$$

where $\mathcal{I}_t^i = \{\theta_t, s_t^i, \overline{\eta}_t^i, p_t\}$ is the information set of an investor i in period t.

The optimality condition for an investor i in period t satisfies

$$q_t^i = \frac{\mathbb{E}\left[\theta_{t+1} + R^{-1}p_{t+1}|\mathcal{I}_t^i\right] - p_t}{\gamma\mathbb{V}ar\left[\theta_{t+1} + R^{-1}p_{t+1}|\mathcal{I}_t^i\right]}.$$

In a stationary equilibrium in linear strategies, the equilibrium demand of investor i can be expressed as

$$\Delta q_t^i = \alpha_\theta^i\theta_t + \alpha_s^i s_{it} + \alpha_\eta^i\overline{\eta}_t^i - \alpha_p^i p_t + \psi^i, \tag{10}$$

where α_θ^i, α_n^i, α_p^i, and ψ^i are individual equilibrium demand coefficients, whose expressions are derived in the Appendix. Market clearing and the Strong Law of Large Numbers allows us to express the equilibrium price in period t as

$$p_t = \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t + \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\eta_t + \frac{\overline{\alpha_\eta}}{\overline{\alpha_p}}n_t + \frac{\overline{\psi}}{\overline{\alpha_p}},$$

and the unbiased signal of the innovation to the fundamental contained in the price can be defined as $\hat{p}_t = \frac{\overline{\alpha_p}}{\overline{\alpha_s}}\left(p_t - \frac{\overline{\alpha_\eta}}{\overline{\alpha_p}}\mu_n - \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t - \frac{\overline{\psi}}{\overline{\alpha_p}}\right)$, so

$$\hat{p}_t = \eta_t + \frac{\overline{\alpha_\eta}}{\overline{\alpha_s}}\left(n_t - \mu_n\right),$$

as in Section 2. Note that $\hat{p}_t|\theta_{t+1}, \theta_t \sim N\left(\eta_t, \tau_{\hat{p}}^{-1}\right)$, with absolute price informativeness given by

$$\tau_{\hat{p}} = \left(\mathbb{V}\mathrm{ar}\left[\hat{p}|\theta_{t+1}, \theta_t\right]\right)^{-1} = \left(\frac{\overline{\alpha_s}}{\overline{\alpha_\eta}}\right)^2\tau_n.$$

Lemma 1. (Assumptions 1 and 2 satisfied) *The set of assumptions considered in Section 2 are such that Assumptions 1 and 2 are endogenously satisfied. Therefore, the model described in this section is a special case of the more general framework studied in Section 2.*

We are ready to introduce the new identification result that allows to recover several primitives of the model in this environment in the following Proposition.

Proposition 3. *(Identifying signal precision and trading needs precision) Consider the environment described in Section 3 and assume that $\tau_s^i = \tau_s$ and $\tau_n^i = \tau_n$, $\forall i$. Let β_0, β_1, and β_2 denote the coefficients of Regression R1, and let ζ_0 and ζ_1 denote the coefficients of Regression R2. The precision* *of investors' private signals τ_s, and the precision of the aggregate component of investors trading needs (noise), τ_n, can be recovered as follows:*

$$\tau_s = \tau_{\hat{p}} \left(\frac{\beta_2}{1 + R^{-1}\zeta_1 - \beta_2} \frac{1}{\tau_{\hat{p}}^R} - 1 \right) \qquad (\textit{Signal Precision})$$

$$\tau_n = \tau_{\hat{p}} \left(\frac{\beta_2}{1 + R^{-1}\zeta_1 - \beta_2} - \tau_{\hat{p}}^R \right)^{-2}, \qquad (\textit{Noise})$$

where $\tau_{\hat{p}}$ and $\tau_{\hat{p}}^R$ are respectively recovered as in Propositions 1 and 2 and R corresponds to the risk-free rate.

Intuitively, in this model, $\beta_2 = \frac{\overline{\alpha_s}}{\overline{\alpha_p}}$ can be interpreted as the share of total ex-post information (measured in precisions) that is acquired by an investor either privately by observing a private signal, or publicly, by conditioning on the price. When β_2 is high, it means that investors priors are relatively unimportant, and that investors mostly trade on private information or acquired public information, which suggests that τ_s is likely to be high and that τ_n is likely to be low. When $\tau_{\hat{p}}$ and $\tau_{\hat{p}}^R$ are high too, this implies that it is likely that investors learn more from the price in relative terms, which suggest that τ_s is likely to be low and that τ_n is likely to be high. A high ζ_1 is associated with finding a low τ_s and a high τ_n, given the other estimates, since it increases the amount of common public information.

To our knowledge, we provide the first approach that enables to directly recover measures of noise trading in Rational Expectations models. However, it's worth highlighting that the identification result in Proposition 3 holds for a much more restrictive set of assumptions than the results in Propositions 1 and 2. In this section, we are taking a particular stance regarding the form of noise trading and we are restricting the information structure to be homogeneous across investors.

4 Extensions

For expositional purposes, we have developed our approach in the context of the single asset model with a stationary payoff process. However, it's worth understanding how to extend our methodology to more general environments. In this section, we extend our results to the case with non-stationary payoffs, multiple risky assets, payoffs with learnable and unlearnable components, and public signals.

4.1 Non-stationary Payoff

It is well known, see e.g., Campbell (2017) for a recent discussion of the literature, that assuming that measures of asset payoffs (dividends, earnings, etc) are non-stationary is often perceived as a better assumption. We show how to adapt our results to an environment in which the asset payoff follows a non-stationary process. Formally, we assume that the asset payoff follows a random walk with drift

$$\Delta\theta_{t+1} = \mu_\theta + \eta_t,$$

which corresponds to our baseline specification setting $\rho = 1$.[11] In this case, Equation (3) remains valid, but the fact that the asset payoff and prices are non-stationary complicates the estimation.[12] However, taking first differences of the equilibrium pricing equation, we get to the following expression in differences

$$\Delta p_t = \left(\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}} - \frac{\overline{\alpha_s}}{\overline{\alpha_p}} \right) \Delta\theta_t + \frac{\overline{\alpha_s}}{\overline{\alpha_p}} \Delta\theta_{t+1} + \frac{\overline{\alpha_n}}{\overline{\alpha_p}} \Delta n_t, \tag{11}$$

in which all variables are stationary. In this case, we can recover absolute price informativeness as follows.

Proposition 4. (Identifying absolute price informativeness, non-stationary payoff) *Assume that the additive noise and the linear asset demands assumptions are satisfied. Let β_0, β_1, and β_2 denote the coefficients of the following regression of prices on fundamentals,*

$$\Delta p_t = \beta_0 + \beta_1 \Delta\theta_t + \beta_2 \Delta\theta_{t+1} + \varepsilon_t, \tag{R4}$$

where p_t denotes the ex-dividend price at the beginning of period t, θ_{t+1} denotes the measure of fundamentals realized over period t, and where we denote the variance of the error by $\sigma_\varepsilon^2 = \mathbb{V}ar[\varepsilon_t]$. Then, absolute price informativeness, $\tau_{\hat{p}}$, can be recovered as follows

$$\tau_{\hat{p}} = 2 \frac{\beta_2^2}{\mathbb{V}ar[\varepsilon_t]}. \tag{12}$$

Intuitively, since we can express $\mathbb{V}\mathrm{ar}[\varepsilon_t]$ in the following way,

$$\mathbb{V}\mathrm{ar}[\varepsilon_t] = \left(\frac{\overline{\alpha_n}}{\overline{\alpha_p}} \right)^2 \mathbb{V}\mathrm{ar}[\Delta n_t] = \left(\frac{\overline{\alpha_n}}{\overline{\alpha_p}} \right)^2 2\mathbb{V}\mathrm{ar}[n_t],$$

we can combine $\mathbb{V}\mathrm{ar}[\varepsilon_t]$ and β_2 to recover $\tau_{\hat{p}}$. As in the case in which payoffs are stationary, we can also recover relative price informativeness without estimating the process for the fundamental by running our regressions in differences, as follows.

Proposition 5. (Identifying relative price informativeness, non-stationary payoff) *Let $R^2_{|\Delta\theta_{t+1},\Delta\theta_t} \equiv 1 - \frac{\mathbb{V}ar[\varepsilon_t]}{\mathbb{V}ar[\Delta p_t]}$ be the R-squared of Regression R4. Let $R^2_{|\Delta\theta_t}$, ζ_0 and ζ_1 respectively denote the R-squared and the coefficients of the following regression R5 of price changes on lagged changes on fundamentals*

$$\Delta p_t = \zeta_0 + \zeta_1 \Delta\theta_t + \varepsilon_t^\zeta. \tag{R5}$$

Then, relative price informativeness, $\tau_{\hat{p}}^R$, can be recovered as follows

$$\tau_{\hat{p}}^R = \frac{\tau_{\hat{p}}}{\tau_\eta} = 2 \frac{R^2_{|\Delta\theta_{t+1},\Delta\theta_t} - R^2_{|\Delta\theta_t}}{1 - R^2_{|\Delta\theta_{t+1},\Delta\theta_t}}. \tag{13}$$

[11] Our methodology can be extended to allow for more general unit root processes.

[12] If the process for the payoff is non-stationary, estimating Regression R1 using OLS is a spurious regression, using the terminology of Granger and Newbold (1974).

Remarkably, the formulas that recover absolute and relative price informativeness when the payoff process corresponds to a random walk are identical to those in the stationary model with two modifications. First, the relevant underlying regressions ought to be run in differences, not in levels to obtain unbiased estimates of the coefficients. Second, the formulas for absolute and relative price informativeness are multiplied by a factor for two. This factor accounts for the fact that the error term in the representation in differences is twice as noisy as the one in the equation in levels, since it is given by the difference of the realizations of n_t and n_{t-1}, which are independent and identically distributed.

4.2 Multiple Risky Assets

We now consider a multi-asset extension to our baseline model and show that an appropriate reinterpretation of aggregate noise allows us to use the single market framework for measurement purposes. The goal of studying this more general framework is two-fold. First, it allows us to reinterpret the results of the single-asset model when many assets are available. Second, it suggests how to use our approach more generally to answer different questions about price informativeness.

We assume that there are N risky assets indexed by $j \in \{1, 2, \ldots, N\}$, with payoffs distributed according to

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\mu}_\theta + C\boldsymbol{\theta}_t + \boldsymbol{\eta}_t, \tag{14}$$

where $\boldsymbol{\mu}_\theta$, $\boldsymbol{\theta}_t$ and $\boldsymbol{\eta}_t$ are $N \times 1$ vectors and C is an $N \times N$ matrix such that the process $\boldsymbol{\theta}_{t+1}$ is stationary.

The counterparts of assumptions 1 and 2 for the baseline environment are as follows.

Assumption 3. (Additive noise) *Each period t, every investor i receives a vector of unbiased private signals s_t^i about the vector of innovations to the payoffs, η_t, of the form*

$$\boldsymbol{s}_t^i = \boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_{st}^i, \tag{15}$$

where $\boldsymbol{\varepsilon}_{st}^i$, $\forall i \in I$, $\forall t$, are vectors of random variables with mean zero and finite second moments, whose realizations are independent across investors and over time. Each period t, every investor i has a vector of private trading needs $\boldsymbol{n}_t^i$, of the form

$$\boldsymbol{n}_t^i = \boldsymbol{n}_t + \boldsymbol{\varepsilon}_{nt}^i, \tag{16}$$

where $\boldsymbol{n}_t$ is a vector of random variables with mean $\boldsymbol{\mu}_n$ and finite second moments, and where $\boldsymbol{\varepsilon}_{nt}^i$, $\forall i \in I$, $\forall t$, are vectors of random variables with mean zero and finite second moments, whose realizations are independent across investors and over time.

Assumption 4. (Linear asset demands) *Investors' net asset demands satisfy*

$$\Delta \boldsymbol{q}_t^i = A_\theta^i \boldsymbol{\theta}_t + A_s^i \boldsymbol{s}_t^i + A_n^i \boldsymbol{n}_t^i - A_p^i \boldsymbol{p}_t + A_0^i, \tag{17}$$

where A_s^i, A_n^i, A_p^i, and A_0^i are $N \times N$ matrices of individual demand coefficients, determined in equilibrium.

Given both Assumptions, market clearing for the risky asset implies that $\int \Delta \boldsymbol{q}_t^i di = 0$, $\forall t$, which, exploiting a Law of Large Numbers, yields a pricing equation of the form

$$\boldsymbol{p}_t = \left(\overline{A_p}\right)^{-1} \overline{A_s} \boldsymbol{\eta}_t + \left(\overline{A_p}\right)^{-1} \overline{A_\theta} \boldsymbol{\theta}_t + \left(\overline{A_p}\right)^{-1} \overline{A_n} \boldsymbol{n}_t + \left(\overline{A_p}\right)^{-1} \overline{A_0}, \tag{18}$$

where we denote the cross sectional averages of individual demand coefficients by $\overline{A_s} = \int_I A_s^i di$, $\overline{A_p} = \int_I A_p^i di$, $\overline{A_\theta} = \int_I A_\theta^i di$, $\overline{A_n} = \int_I A_n^i di$, and $\overline{A_0} = \int A_0^i di$.

In a multi-asset environment, there are several notions of price informativeness. The vector of prices $\hat{\boldsymbol{p}}_t$ contains information about the vector of fundamentals $\boldsymbol{\theta}_{t+1}$. At the same time, each price p_j contains information about the vector of fundamentals. Since we are interested in the precision of the price $p_{j,t}$ as a signal of the fundamental $\theta_{j,t+1}$, our notion of absolute price informativeness along asset dimension j is given by

$$\tau_{\hat{p}}^j = \mathbb{V}\mathrm{ar}\left[\hat{p}_j | \theta_{j,t+1}, \theta_{j,t}\right]^{-1},$$

where $\hat{p}_j$ is the unbiased signal about $\theta_{j,t+1}$ contained in the price. $\tau_{\hat{p}}^j$ is the precision of the unbiased signal contained about the fundamental of asset j from the perspective of an external observer who sees the previous realization of the fundamental of asset j only. The price of asset j is given by

$$p_{j,t} = \left[\left(\overline{A_p}\right)^{-1} \overline{A_0}\right]_j + \sum_h \left(\pi_{jh} \left[\left(\overline{A_p}\right)^{-1} \overline{A_s}\right]_{jh}\right) \theta_{j,t+1} + \sum_h \left(\pi_{jh} \left[\left(\overline{A_p}\right)^{-1} \left(\overline{A_\theta} - \overline{A_s} C\right)\right]_{jh}\right) \theta_{j,t} + \Gamma_j \boldsymbol{u}_{j,t}$$

where $\boldsymbol{u}_{j,t}$ is the vector of all the trading motives of investors that are orthogonal to the fundamentals of asset j and Γ_j is a function of aggregate equilibrium demand sensitivities. More specifically

$$\boldsymbol{u}_{j,t} = \left[\boldsymbol{\omega}_t^{j\prime}, \boldsymbol{\omega}_{t-1}^{j\prime}, \boldsymbol{n}_t'\right]',$$

where $\boldsymbol{\omega}_t^j = \left[\omega_{1t}^j, ..., \omega_{Nt}^j\right]$ is orthogonal to $\theta_{j,t+1}$ and $\omega_{h,t}^j$ is defined as the residual of a regression of $\theta_{h,t+1}$ on $\theta_{j,t+1}$, i.e.,

$$\omega_{h,t}^j = \theta_{h,t+1} - \frac{\mathbb{C}\mathrm{ov}\left[\theta_{h,t+1}, \theta_{j,t+1}\right]}{\mathbb{V}ar\left[\theta_{h,t+1}\right]} \theta_{j,t+1}.$$

Hence, $\omega_{j,t}^j = 0$. Moreover,

$$\Gamma_j = \left[\left[\left(\overline{A_p}\right)^{-1} \overline{A_s}\right]_j, \left[\left(\overline{A_p}\right)^{-1} \left(\overline{A_\theta} - \overline{A_s} C\right)\right]_j, \left[\left[\left(\overline{A_p}\right)^{-1} \overline{A_n}\right]_j\right]\right].$$

If the realizations of $\theta_{h,t+1}$ and $\theta_{j,t+1}$ are uncorrelated, then the price will put weight 0 on $\omega_{h,t}^j$.

From the perspective of an external observer who only observes the current fundamental of asset j, the unbiased signal about the future fundamental of asset j contained in the price is

$$\begin{aligned}\hat{p}_{j,t} &= \left(\sum_h \left(\pi_{jh} \left[\left(\overline{A_p}\right)^{-1} \overline{A_s}\right]_{jh}\right)\right)^{-1} \left(p_{j,t} - \left[\left(\overline{A_p}\right)^{-1} \overline{A_0}\right]_j - \sum_h \left(\pi_{jh} \left[\left(\overline{A_p}\right)^{-1} \left(\overline{A_\theta} - \overline{A_s} C\right)\right]_{jh}\right) \theta_{j,t}\right) \\ &= \theta_{j,t+1} + \left(\sum_h \left(\pi_{jh} \left[\left(\overline{A_p}\right)^{-1} \overline{A_s}\right]_{jh}\right)\right)^{-1} \Gamma_j \boldsymbol{u}_{j,t}.\end{aligned}$$

The precision of this signal is our measure of price informativeness and it is given by

$$\tau_{\hat{p}}^j = \mathbb{V}\mathrm{ar}\left[\hat{p}_{j,t} | \theta_{j,t+1}, \theta_{j,t}\right]^{-1} = \left(\sum_h \left(\pi_{jh} \left[\left(\overline{A_p}\right)^{-1} \overline{A_s}\right]_{jh}\right)\right)^2 \mathbb{V}\mathrm{ar}\left[\Gamma_j \boldsymbol{u}_{j,t}\right]^{-1}.$$

Proposition 6. (Identifying relative price informativeness through univariate regression) *Assume that the additive noise and the linear asset demands assumptions are satisfied. Let β_0, β_1, and β_2 denote the coefficients of the following regression of prices on fundamentals. This measure of price informativeness can be recovered from the asset specific regression*

$$p_{j,t} = \beta_0 + \beta_1 \theta_{j,t} + \beta_2 \theta_{j,t+1} + \varepsilon_{j,t}$$

where p_t denotes the ex-dividend price at the beginning of period t, $\theta_{j,t+1}$ denotes the measure of fundamentals for asset j realized over period t, and where we denote the variance of the error by $\sigma^2_{\varepsilon j} = \mathbb{V}ar\left[\varepsilon_{j,t}\right]$. Then, absolute price informativeness, $\tau^j_{\hat{p}}$, can be recovered as follows

$$\tau^j_{\hat{p}} = \frac{\beta_2^2}{\sigma^2_{\varepsilon j}}.$$

The main difference between the single asset and the multi-asset version of our model is the interpretation of the noise that prevents the price from being fully revealing. In the single asset case, the noise is given purely by the aggregate trading motives, which is orthogonal to the innovation to the fundamental. In the multi-asset case, the noise is a combination of the aggregate trading motives for all assets in the economy and the components in the realizations of the fundamentals of other assets that are orthogonal to the fundamental of the asset in which one is interested. These additional sources of noise enter the pricing equation when the fundamentals are correlated across assets. In this case, a signal about the innovation to the fundamental of one asset can also be used to learn about the innovation to the fundamental of another asset. In our general approach in Section 2 we are not restricting the source of the noise. Hence, the analysis in the previous sections accommodates correlated asset payoffs in multi-asset environments.

4.3 Learnable and Unlearnable Payoff

Our results so far imply that if investors could fully aggregate their dispersed information, they would be able to fully learn the asset payoff. In this subsection, we consider the possibility that part of the asset payoff is simply unlearnable for investors at the trading stage. Formally, we assume that the innovation to the asset payoff has learnable and unlearnable components, so the asset payoff can be written as

$$\theta_{t+1} = \rho \theta_t + \eta_t \quad \text{where} \quad \eta_t = \eta^L_t + \eta^U_t,$$

where the unlearnable component η^U_t is random, has mean zero and finite variance, and its realization are independently distributed from other random variables. Investors exclusively receive signals about the learnable component of the asset payoff, so formally

$$s^i_t = \eta^L_t + \varepsilon^i_{st}.$$

Investors' signals can be reformulated as follows

$$s^i_t = \eta_t - \eta^U_t + \varepsilon^i_{st} = \eta_t + \varepsilon^{i\prime}_{st},$$

where $\varepsilon_{st}^{i\prime} = -\eta_t^U + \varepsilon_{st}^i$, which allows to derive the following expression for the equilibrium price

$$p_t = \frac{\overline{\psi}}{\overline{\alpha_p}} + \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t + \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\eta_t + \underbrace{\frac{\overline{\alpha_n}}{\overline{\alpha_p}}n_t - \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\eta_t^U}_{\varepsilon_t},$$

where the unbiased signal about the innovation to the fundamental corresponds to $\hat{p}_t = \frac{\overline{\alpha_p}}{\overline{\alpha_s}}p_t - \frac{\overline{\alpha_\theta}}{\overline{\alpha_s}}\theta_t - \frac{\overline{\psi}}{\overline{\alpha_s}} = \eta_t + \frac{\overline{\alpha_n}}{\overline{\alpha_s}}n_t - \eta_t^U$.

Proposition 7. (Identifying Price Informativeness with Learnable and Unlearnable Payoffs) *Assume that the additive noise and the linear asset demands assumptions are satisfied. Then, absolute and relative price informativeness can be recovered from Regressions R1 and R2 as follows*

$$\tau_{\hat{p}} = \frac{\beta_2^2}{\sigma_\varepsilon^2} \quad \textit{and} \quad \tau_{\hat{p}}^R = \frac{R^2_{|\theta_{t+1},\theta_t} - R^2_{|\theta_t}}{1 - R^2_{|\theta_{t+1},\theta_t}}.$$

It should be evident that the identification results from Propositions 1 and 2 apply to this case, provided that we reinterpret the noise component by including the uncertainty about the unlearnable component of the asset payoff. Conceptually, the fact that the information received by investors as a whole is not enough to recover the asset payoff means that prices have to be less informative.

4.4 Public Signals

In our results until now, we have considered private signals as the only source of information in the economy. In this subsection, we consider the case in which investors also observe a public signal about the fundamental. We extend the environment in the baseline model by allowing investors to observe a public signal with the following structure

$$\pi_t = \eta_t + \varepsilon_{\pi t}$$

where $\varepsilon_{\pi t} \sim N\left(0, \tau_\pi^{-1}\right)$ is iid across time and independent of the innovations η_t. In this case, we augment the linear asset demands considered in Assumption 2 as follows.

Assumption 5. (Linear asset demands with public signals) *Investors' net asset demands satisfy*

$$\Delta q_t^i = \alpha_\theta^i \theta_t + \alpha_s^i s_t^i + \alpha_\pi^i \pi_t + \alpha_n^i n_t^i - \alpha_p^i p_t + \psi^i,$$

where α_s^i, α_θ^i, α_n^i, α_π^i, α_p^i, and ψ^i are individual demand coefficients, determined in equilibrium.

Given our assumptions, market clearing in the risky asset market implies

$$p_t = \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t + \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\eta_t + \frac{\overline{\alpha_\pi}}{\overline{\alpha_p}}\pi_t + \frac{\overline{\alpha_n}}{\overline{\alpha_p}}n_t + \frac{\overline{\psi}}{\overline{\alpha_p}},$$

where we denote the cross sectional averages of individual demand coefficients by $\overline{\alpha_s} = \int_I \alpha_s^i di$, $\overline{\alpha_p} = \int_I \alpha_p^i di$, $\overline{\alpha_\theta} = \int_I \alpha_\theta^i di$, $\overline{\alpha_n} = \int_I \alpha_n^i di$, $\overline{\alpha_\pi} = \int_I \alpha_\pi^i di$, and $\overline{\psi} = \int_I \psi^i di$.

The unbiased signal contained in the price from the perspective of an external investor depends on his information set. If the external investor only observes the price and the past realization of the fundamental but not the public signal, the unbiased signal contained in the price is given by

$$\begin{aligned}\hat{p}_t &= \frac{\overline{\alpha_p}}{\overline{\alpha_s}+\overline{\alpha_\pi}}\left(p_t - \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t - \frac{\overline{\alpha_n}}{\overline{\alpha_p}}\mu_n - \frac{\overline{\psi}}{\overline{\alpha_p}}\right)\\ &= \eta_t + \frac{\overline{\alpha_\pi}}{\overline{\alpha_s}+\overline{\alpha_\pi}}\varepsilon_{\pi t} + \frac{\overline{\alpha_n}}{\overline{\alpha_s}+\overline{\alpha_\pi}}\left(n_t - \mu_n\right)\end{aligned}$$

and absolute price informativeness is given by

$$\tau_{\hat{p}} \equiv \left(\mathbb{V}\mathrm{ar}\left[\hat{p}_t|\theta_{t+1},\theta_t\right]\right)^{-1} = \frac{\left(\overline{\alpha_s}+\overline{\alpha_\pi}\right)^2}{\left(\overline{\alpha_\pi}\right)^2\tau_\pi^{-1} + \left(\overline{\alpha_n}\right)^2\tau_n^{-1}}. \tag{19}$$

Proposition 8. (Identifying price informativeness with public signals) *Assume that the additive noise and the linear asset demands assumptions are satisfied. Then, absolute and relative price informativeness can be recovered from Regressions R1 and R2 as follows*

$$\tau_{\hat{p}} = \frac{\beta_2^2}{\sigma_\varepsilon^2} \quad \textit{and} \quad \tau_{\hat{p}}^R = \frac{R^2_{|\theta_{t+1},\theta_t} - R^2_{|\theta_t}}{1-R^2_{|\theta_{t+1},\theta_t}}.$$

The first identification result when investors do not observe the public signal becomes evident rewriting the equilibrium price as follows

$$\begin{aligned}p_t &= \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t + \frac{\overline{\alpha_s}+\overline{\alpha_\pi}}{\overline{\alpha_p}}\eta_t + \frac{\overline{\alpha_n}}{\overline{\alpha_p}}n_t + \frac{\overline{\alpha_\pi}}{\overline{\alpha_p}}\varepsilon_{\pi t} + \frac{\overline{\psi}}{\overline{\alpha_p}}\\ &= \left(\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}} - \rho\frac{\overline{\alpha_s}+\overline{\alpha_\pi}}{\overline{\alpha_p}}\right)\theta_t + \underbrace{\frac{\overline{\alpha_s}+\overline{\alpha_\pi}}{\overline{\alpha_p}}}_{=\beta_2}\theta_{t+1} + \underbrace{\frac{\overline{\alpha_n}}{\overline{\alpha_p}}n_t + \frac{\overline{\alpha_\pi}}{\overline{\alpha_p}}\varepsilon_{\pi t}}_{=\varepsilon_t} + \frac{\overline{\psi}}{\overline{\alpha_p}}.\end{aligned} \tag{20}$$

It is obvious from Eq. (20) that the coefficient β_2 in regression R1 recovers

$$\beta_2 = \frac{\overline{\alpha_s}+\overline{\alpha_\pi}}{\overline{\alpha_p}}$$

and the error term maps to $\varepsilon_t = \frac{\overline{\alpha_n}}{\overline{\alpha_p}}n_t + \frac{\overline{\alpha_\pi}}{\overline{\alpha_p}}\varepsilon_{\pi t}$. This implies that our estimates of absolute price informativeness derived in the baseline model would exactly recover price informativeness from the perspective of an external observer who only observes the price and has no access to the private signal π since

$$\frac{\beta_2^2}{\mathbb{V}\mathrm{ar}\left[\varepsilon_t\right]} = \frac{\left(\overline{\alpha_s}+\overline{\alpha_\pi}\right)^2}{\left(\overline{\alpha_n}\right)^2\tau_n^{-1} + \left(\overline{\alpha_\pi}\right)^2\tau_\pi^{-1}},$$

which is exactly price informativeness in Eq. (19). The second result identifying relative price informativeness when the external observer does not observe the public signal available to the investors follows exactly from the proof in the baseline model reinterpreting the loading on the fundamental and the noise term to include the information and noise in the public signal. Proposition 8 implies that our baseline estimates or absolute and relative price informativeness are valid when investors have access to common information that the external observer does not see. Having omitted public signals in our regressions is not a concern provided that we define price informativeness appropriately.

So far, we have limited the information available to the external observer from whose perspective we are defining price informativeness. However, there many situations in which the nature of the public signal is such that even an external observer has access to it. In this case, the relevant concept of informativeness should be modified to condition on the additional information that is available to the external observer. More specifically, if the external observer has access to the public signal as well as the price and the past realization of the asset payoff, the unbiased signal about the fundamental contained in the price is given by

$$\begin{aligned}\hat{p}_t &= \frac{\overline{\alpha_p}}{\overline{\alpha_s}}\left(p_t - \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t - \frac{\overline{\alpha_\pi}}{\overline{\alpha_p}}\pi_t - \frac{\overline{\alpha_n}}{\overline{\alpha_p}}\mu_n - \frac{\overline{\psi}}{\overline{\alpha_p}}\right)\\ &= \eta_t + \frac{\overline{\alpha_n}}{\overline{\alpha_s}}\left(n_t - \mu_n\right)\end{aligned}$$

and the precision of this signal is given by

$$\tilde{\tau}_{\hat{p}} \equiv \left(\mathbb{V}\mathrm{ar}\left[\hat{p}_t|\theta_{t+1}, \theta_t, \pi_t\right]\right)^{-1} = \left(\frac{\overline{\alpha_s}}{\overline{\alpha_n}}\right)^2 \tau_n. \tag{21}$$

We refer to the measure of informativeness in Eq. (21) as "informed" absolute price informativeness. As in the baseline model, we define "informed" relative price informativeness as $\tilde{\tau}^R_{\hat{p}} = \frac{\tilde{\tau}_{\hat{p}}}{\tau_\eta}$.

Proposition 9. (Identifying "informed" price informativeness with public signals) *Assume that the additive noise and the linear asset demands assumptions are satisfied. Then, "informed" absolute and relative price informativeness can be recovered from regressions of prices on fundamentals and the public signal*

$$p_t = \chi_0 + \chi_1\theta_t + \chi_2\theta_{t+1} + \chi_3\pi_t + \tilde{\varepsilon}_t, \tag{R6}$$

and

$$p_t = \tilde{\chi}_0 + \tilde{\chi}_1\theta_t + \tilde{\chi}_2\theta_{t+1} + \varepsilon^{\chi}_t, \tag{R7}$$

as

$$\tilde{\tau}_{\hat{p}} = \frac{\chi_2^2}{\sigma^2_{\tilde{\varepsilon}}} \quad \text{and} \quad \tilde{\tau}^R_{\hat{p}} = \frac{\tilde{R}^2_{|\theta_{t+1},\theta_t} - \tilde{R}^2_{|\theta_t}}{1 - \tilde{R}^2_{|\theta_{t+1},\theta_t}},$$

where $\tilde{R}^2_{|\theta_{t+1},\theta_t}$ and $\tilde{R}^2_{|\theta_t}$ are the R-squareds of Regression R6 and Regression R7, respectively.

The results in the proposition above follow directly by noting that $\chi_2 = \frac{\overline{\alpha_s}}{\overline{\alpha_p}}$ and $\tilde{\varepsilon}_t = \frac{\overline{\alpha_n}}{\overline{\alpha_p}}n_t$ and using the definition of "informed" absolute price informativeness in Eq. (21). The intuition behind the results for relative price informativeness follows exactly the same logic as in the baseline model.

5 Empirical Implementation

In this section, we empirically implement our identification strategy and recover stock-specific measures of price informativeness. In the text, we provide a brief description of the data and the sample selection

Table 1: Summary Statistics (All Observations)

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Market Cap.	88,356	3,336.34	17,681.97	0.72	81.45	352.77	1,532.62	682,084.00
Earnings	88,356	93.14	595.14	−15,181.20	0.71	7.78	39.43	33,304.45

Note: Table 1 presents summary statistics for the full sample of $88,256$ stock-year observations. It provides information on the sample mean, median, and standard deviation, as well as the minimum, the maximum and the 25th and 75th percentiles of the distribution of market capitalization and total earnings. All variables are expressed in millions of dollars in 2008.

procedure, leaving a more detailed description to the companion R notebooks. Our theoretical results show that running cross-sectional regressions to recover measures of price informativeness imposes strong assumptions about primitives across stocks. In particular, it imposes that the fundamental has the same volatility for all stocks in the sample at any point in time. Consistent with the identification results presented in our theoretical analysis, we instead run time-series regressions at the stock level to recover estimates of absolute price informativeness, relative price informativeness, and Kalman gains.

5.1 Data Description

We now describe the data sources and variables used to implement the methodology we develop in the previous sections to obtain measures of price informativeness. We conduct our analysis using data from 1963 to 2017. We obtain stock price data from the Center for Research in Security Prices (CRSP) to calculate stocks market values, data on reported earnings, to use as a measure of fundamentals, from CRSP/Compustat Merged (CCM), and a personal consumption deflator index from FRED.

In this section, we use a sample with all CRSP common stocks at a quarterly frequency. In the Appendix, we report the results of our analysis at an annual frequency. Without loss of generality, to avoid dealing with float issues, we use market value, which we denote by M, as the relevant measure for the value/price of a firm. We use firms' total earnings, as measured by EBIT, which we denote by E, as the relevant measure of firms' payoffs.[13] To match the timing of our model and ensure that the realized earnings are observed at the time at which the price is determined, we match the date t EBIT, E_t, with the price one quarter forward.[14] For example, EBIT on March 1986 corresponds to θ_t in our model and the stock price on June 1986 corresponds to p_t. Similarly, the realization of the fundamental at the end of period t is given by the reported EBIT a quarter later. The realized payoff on March 1986 is given by EBIT on June 1986, which corresponds to θ_{t+1} in our model. Finally, we

[13] The choice of payoff measures may be objectionable. Dividend measures at the stock level are noisy. As it is customary in the literature that studies individual stocks, e.g. Vuolteenaho (2002), we map payoffs in the model to measures of earnings.

[14] For our analysis at a quarterly frequency, we use the price one month forward. Using contemporary variables yields similar results, which are available in the R code provided online.

Table 2: Summary Statistics (Mean and Standard Deviation of Earnings)

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Mean Earnings	839	77.04	344.25	−104.39	1.21	11.76	43.69	5,731.18
St. Dev. Earnings	839	81.06	371.29	0.18	3.41	12.23	39.49	5,693.84

Note: Table 2 presents summary statistics for the full sample of 839 stocks. It provides information on the sample mean, median, and standard deviation, as well as the minimum, the maximum and the 25th and 75th percentiles of the distribution of the variance of earnings. All variables are expressed in millions of dollars in 2008.

run our regressions at the stock level and include only firms for which we have at least 80 observations available.

Table 1 shows summary statistics for our full sample of $88,356$ stock-quarter observations. Our sample exhibits considerable variation in terms of market capitalization and total earnings. The distribution of market capitalization across firms and periods has a mean of $\$3,336$ million, a median of \$352 million and a standard deviation of $\$17,681$ million. The minimum market capitalization in a given quarter is \$0.72 million and the maximum is $\$682,962$ million. The distribution of total earnings across firms and periods has a mean of \$93 million, a median of \$7.8 million, and a standard deviation of \$595 million.

Table 2 shows summary statistics at the stock level for the 839 stocks with more than 80 observations. In particular, this table summarizes the differences in the distribution of earnings across stocks. The mean earnings across stocks have a mean of \$77 million, a median of \$11.8 million and a standard deviation of \$344 million. The median standard deviation in earnings is \$81 million and it exhibits a standard deviation of \$371 million. These summary statistics show that there is significant heterogeneity in the earnings process in the cross-section of firms as the mean and, more importantly, the volatility of the fundamental vary considerably across stocks. This finding questions the validity of using cross-sectional regressions. Figure 2 shows the distribution of standard deviations for earnings across stocks.

5.2 Empirical Specification

We implement Proposition 1 by running time-series regression for each individual stock. On the right-hand side, we have measures of market value, M_t^j. On the left-hand side, we use current earnings, $E_{j,t}$ and earnings one period ahead, $E_{j,t+1}$. Formally, for each stock, which we index by j, we run a time-series regression of the form

$$M_t^j = \beta_0^j + \beta_1^j E_{j,t} + \beta_2^j E_{j,t+1} + \varepsilon_t^j \Rightarrow R^{2j}_{|\theta_{t+1},\theta_t} \tag{22}$$

$$M_t^j = \zeta_0^j + \zeta_1^j E_{j,t} \qquad\qquad + \hat{\varepsilon}_t^j \Rightarrow R^{2j}_{|\theta_t}, \tag{23}$$

where t corresponds to the time index. Variables β_0^j, β_1^j, and β_2^j are the coefficients for each of the regressions, whereas $\varepsilon_{j,t}$ are the error terms. We respectively denote the R-squareds of the regressions

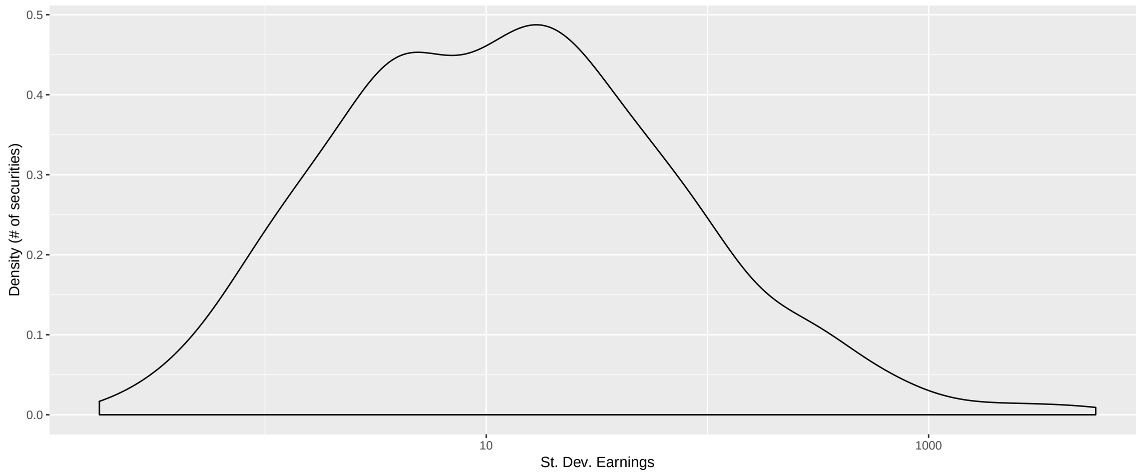

Figure 2: Distribution of standard deviation of total earnings

Note: Figure 2 depicts the distribution of the standard distribution of earnings across the 839 stocks in our sample.

(22) and (23) by $R^{2j}_{|\theta_{t+1},\theta_t}$ and the $R^{2j}_{|\theta_t}$. Hence, Regression R1 in the paper maps to Equation (22), while Regression R2 maps to Equation (23). Using the results in Propositions (1) and (2) we recover absolute and relative price informativeness as follows

$$\tau^j_{\hat{p}} = \frac{\left(\beta^j_2\right)^2}{\mathbb{V}\text{ar}\left[\varepsilon^j_t\right]} \quad \text{and} \quad \tau^{Rj}_{\hat{p}} = \frac{R^{2j}_{|\theta_{t+1},\theta_t} - R^{2j}_{|\theta_t}}{1 - R^{2j}_{|\theta_{t+1},\theta_t}}. \tag{24}$$

One of the main assumptions behind the validity of our methodology running our linear regressions in levels, is the stationarity of the process for earnings. To evaluate the plausibility of this assumption in our data, we run Dickey-Fuller (DF) tests for each stock. For 309 out of 839 stocks in our sample, there is not enough evidence to reject the null hypothesis that a unit root is present in the autoregressive process for earnings.[15] For these stocks, we estimate our measures of price informativeness using the results derived in Section 4.1. We run the following specification in differences

$$\Delta M^j_t = \beta^j_0 + \beta^j_1 \Delta E_{j,t} + \beta^j_2 \Delta E_{j,t+1} + \varepsilon^j_{\Delta t} \qquad \Rightarrow R^{2j}_{|\Delta\theta_{t+1},\Delta\theta_t} \tag{25}$$

$$\Delta M^j_t = \zeta^j_0 + \zeta^j_1 \Delta E_{j,t} \qquad\qquad\qquad + \hat{\varepsilon}^j_{\Delta t} \qquad \Rightarrow R^{2j}_{|\theta_t}, \tag{26}$$

where, analogous to Equation (22) and (23), t and j correspond to the time and stock index, respectively. Equations (25) and (26) respectively map to the regressions in Equations (R4) and (R5), and $R^{2j}_{|\Delta\theta_{t+1},\Delta\theta_t}$and $R^{2j}_{|\Delta\theta_t}$ are their corresponding R-squareds. Hence, using the results in Equations (12) and (13) we recover absolute and relative price informativeness for each stock j as follows

$$\tau^j_{\hat{p}} = 2\frac{\left(\beta^j_2\right)^2}{\mathbb{V}\text{ar}\left[\varepsilon^j_{\Delta t}\right]} \quad \text{and} \quad \tau^{Rj}_{\hat{p}} = 2\frac{R^{2j}_{|\Delta\theta_{t+1},\Delta\theta_t} - R^{2j}_{|\Delta\theta_t}}{1 - R^{2j}_{|\Delta\theta_{t+1},\Delta\theta_t}}. \tag{27}$$

[15] The 839 stocks in our sample are the remaining stocks after removing the stocks with unit roots in the process for ΔE_t.

Finally, we exclude outliers from our sample. We remove all stocks for which the maximum leverage score of any observation is above 0.4 when estimating either Equation (22) or Equation (23), or for their counterparts in differences for those stocks that fail our stationarity tests.

5.3 Empirical Findings

After removing outliers, we have 666 individual stocks with more than 80 quarterly observations. For all stocks that pass our stationarity tests, we recover stock specific measures of absolute and relative price informativeness from Equation (24). For those stocks that fail our stationarity tests, we estimate our measures of price informativeness from Equation (27), after running the difference specifications in Equations (25) and (26). Table 3 shows summary statistics for the distribution of our estimates of stock-specific measures of price informativeness. In the Appendix, we report estimates of our measures of price informativeness assuming the earnings process is non-stationary for all stocks.

We find that, in our sample, the mean absolute price informativeness is 0.04 and the median is 0.001. More importantly, there is significant variation in our estimates of absolute price informativeness in the data. The standard deviation of absolute price informativeness is 0.11, which reinforces our prior about the importance of providing stock-specific measures of informativeness. However, looking at absolute price informativeness may not be the most adequate measure to understand the informational content of prices in the cross section since there are differences in the uncertainty about the fundamental across stocks.

Relative price informativeness normalizes the precision of the signal contained in the price by the volatility of the fundamental for each stock. This normalization makes the comparison across stocks meaningful and more natural to interpret. In our sample, the mean relative price informativeness is around 0.07 which implies that, on average, the precision of the price as a signal of the fundamental is 7% of the precision of the prior. In terms of variances, this translates to the signal contained in the price being 14 times more uncertain than the fundamental. As one would expect, there is a significant variation in relative price informativeness across stocks. The distribution of relative price informativeness is right skewed, with 75% of stocks featuring relative prices informativeness to be less than 8% of the precision of the innovation to earnings.

Though relative price informativeness provides a better context than absolute price informativeness to interpret the magnitudes of the informational content of prices, it is still somewhat difficult to give these numbers an economic interpretation. The best measure of price informativeness to do so is the Kalman gain.[16] There is a one-to-one mapping between the Kalman gain and relative price informativeness given by

$$K_j = \frac{\tau_{\hat{p}}^{Rj}}{1 + \tau_{\hat{p}}^{Rj}}. \tag{28}$$

This expression is bounded between 0 and 1, and it measures the weight that a Bayesian investor puts on the new information revealed by the price when updating his beliefs about the fundamental. For

[16] The economic interpretation of the expression of the Kalman gain in Equation assumes Bayesian updating and Gaussian uncertainty.

example, a Kalman gain of 0.4 implies that a Bayesian investor will put 40% weight on the information contained in the price and 60% on his prior in forming his beliefs. In the limit, when prices are fully revealing, the Kalman gain is equal to 1, and it is 0 when prices contain no information. We focus the rest of our analysis using the Kalman gain as a measure of informativeness.

Table 3: Summary Statistics (Recovered Informativeness Measures)

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Absolute Price Informativeness	666	0.04	0.19	0.00	0.0001	0.001	0.01	2.26
Relative Price Informativeness	666	0.07	0.11	0.00	0.01	0.02	0.08	1.00
Kalman Gain	666	0.05	0.08	0.00	0.01	0.02	0.08	0.50

Note: Table 3 presents summary statistics on the informative measures recovered. It provides information on the sample mean, median, and standard deviation, as well as the 25th and 75th percentiles of the distribution.

The mean Kalman gain in our sample is 0.05 and the median is 0.08. As with our measures of absolute and relative price informativeness, Table 3 shows there is a substantial dispersion in the distribution of Kalman gains across stocks. Prices contain little to no information for 75% of the stocks in our sample, which is reflected by a Kalman gain of less than 0.08. Figure 3 shows the distribution of Kalman gains for all stocks in our sample.

Our approach does not allow us to identify the source of noise impounded in the price. However, our estimates shed light on the amount of noise in financial markets. Given the small magnitude of our estimates of price informativeness, our empirical results suggest that although prices contain some information about the fundamental, they also reflect a considerable amount of noise.

To better understand of the distribution of Kalman gains across stocks, we look at different stock characteristics and their correlation with relative price informativeness. We compare price informativeness in different exchanges, across market capitalization, across the level of volume traded.

An important takeaway from our empirical analysis is the ubiquitous heterogeneity in price informativeness across stocks. This heterogeneity in our estimates across stock characteristics challenges the interpretation of results from cross-sectional regressions as measures of price informativeness, even when controlling for firm-specific characteristics.

Note that our empirical approach uses time series regressions to identify asset-specific measures of price informativeness. This approach contrasts with some of the recent literature that seeks to measure price informativeness from the cross-section of securities, which implicitly imposes that all primitives are constant across asset classes. An advantage of running time-series regressions is that they sidestep potential composition issues, although they introduce potential concerns regarding time variation in the model parameters, for instance, associated with time-varying risk premia or time-varying characteristics.

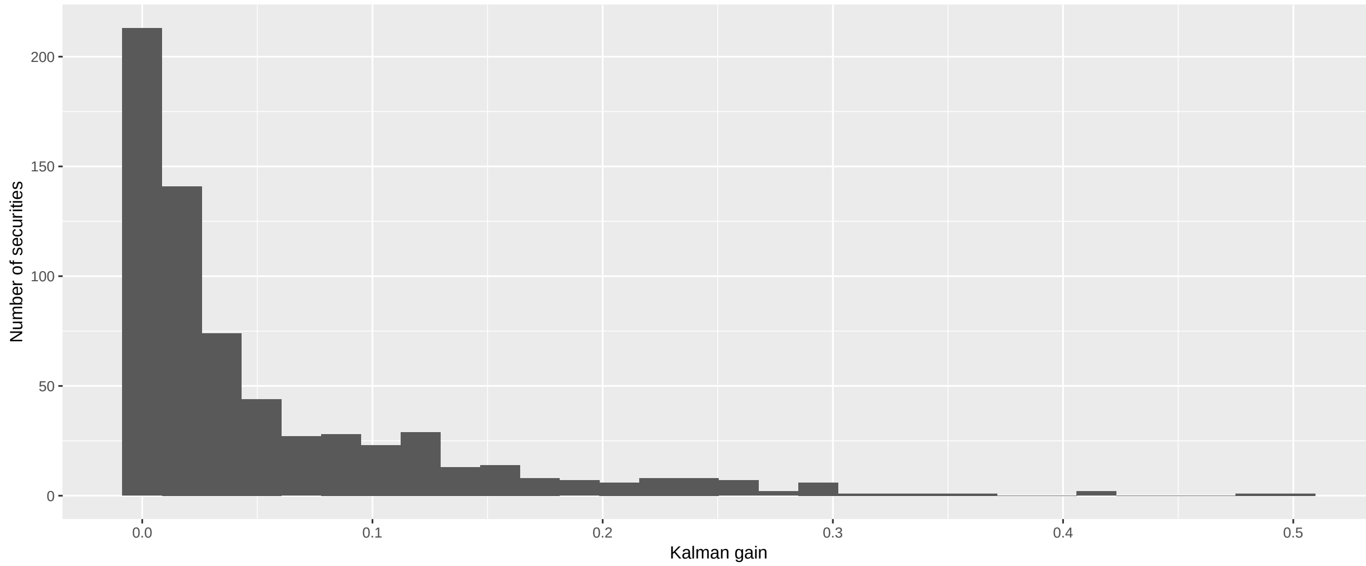

Figure 3: Distribution of Kalman gains

Note: Figure 3 shows the histogram of Kalman gains for the final sample of 666 stocks. The estimates are computed using quarterly data.

5.3.1 Informativeness across exchanges

Prices aggregate the information dispersed in the market. The quality of this aggregation depends on how much information is held by investors and on how much noise is present in individual demands. Markets in which investors have very precise private information about the fundamental and trade mostly on this information will have prices with high price informativeness. Markets in which investors trade primarily on trading motives that are orthogonal to the fundamental will have low price informativeness. Therefore, an important determinant of how much information is contained in asset prices is the set of market participants.

In our sample, stocks are traded in three different exchanges: NYSE, AMEX, and NASDAQ. Figure 4 shows the distribution of Kalman gains for stocks traded in each of these exchanges. We find that, on average, stocks traded in the NYSE have the highest price informativeness, followed by those traded in the NASDAQ. Stocks traded in the AMEX have the lowest level of price informativeness and the lowest dispersion of Kalman gains across the three exchanges.

5.3.2 Informativeness and market capitalization

We look at the relation between price informativeness and market capitalization. A priori, it is not clear what the sign of this correlation should be. For example, on the one hand, one could argue that investors acquire more information about stocks with higher market capitalization because it increases the scale of the trades through which the investor can benefit from his private information. This would imply a positive correlation between price informativeness and market capitalization. On the other hand, bigger firms can also attract more speculative trades and, hence, its stocks can trade at

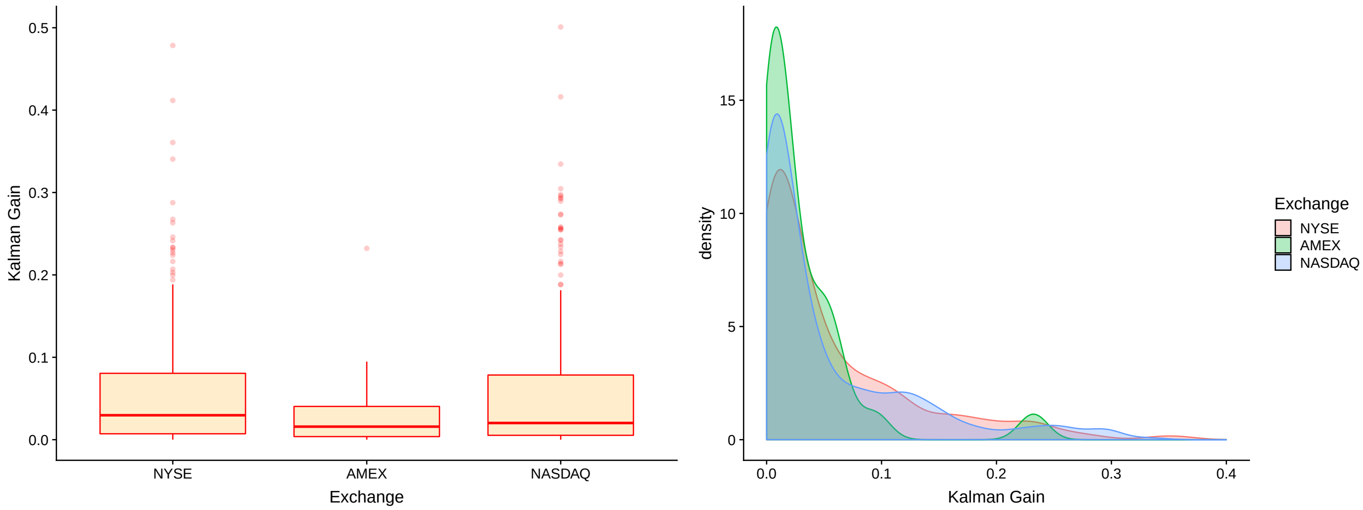

Figure 4: Distribution of Kalman gains across exchanges in which the asset is traded

Note: Figure 4 shows the histogram of Kalman gains for the final sample of 666 stocks sorted by the exchange in which they trade. The estimates are computed using quarterly data.

prices that incorporate a lot of noise and have little informational content. Not surprisingly, whether price informativeness is positively or negatively correlated with market capitalization depends on how the ratio of information to noise impounded in the price varies with this variable. In our sample, we find that the relative price informativeness increases with market capitalization. Figure 5 shows the distribution of Kalman gains across stocks with different average market capitalization. This figure presents a positive relation between our estimated Kalman gains and the average market capitalization of a stock. At the same time, this figure exhibits considerable variation in Kalman gains for a given level of market capitalization.

5.3.3 Informativeness and traded volume

Next, we explore how price informativeness correlates with a stock's traded volume. As in the case of market capitalization, it is not clear what the correlation between traded volume and price informativeness should be. One could argue that high traded volume reflects a large amount of speculative trades and, hence, expect price informativeness and the Kalman gain to be low. Alternatively, high traded volume can reflect large trades by investors with precise information, which would lead to high price informativeness and a high Kalman gain.

Figure 6 shows our estimates of Kalman gains as a function of the stock's average traded volume computed in US dollars. The figure shows a positive correlation between price informativeness and average traded volume. This figure also shows that there is substantial dispersion in Kalman gains after conditioning for the average traded volume.

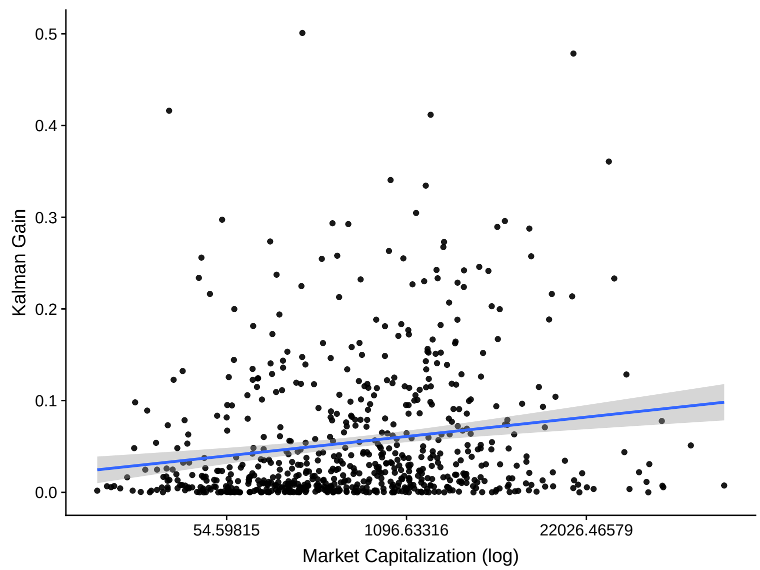

Figure 5: Distribution of Kalman gains across market capitalization

Note: Figure 5 shows the histogram of Kalman gains for the final sample of 666 stocks sorted by their average market capitalization in our sample. The estimates are computed using quarterly data. The blue solid line represents the linear regression line.

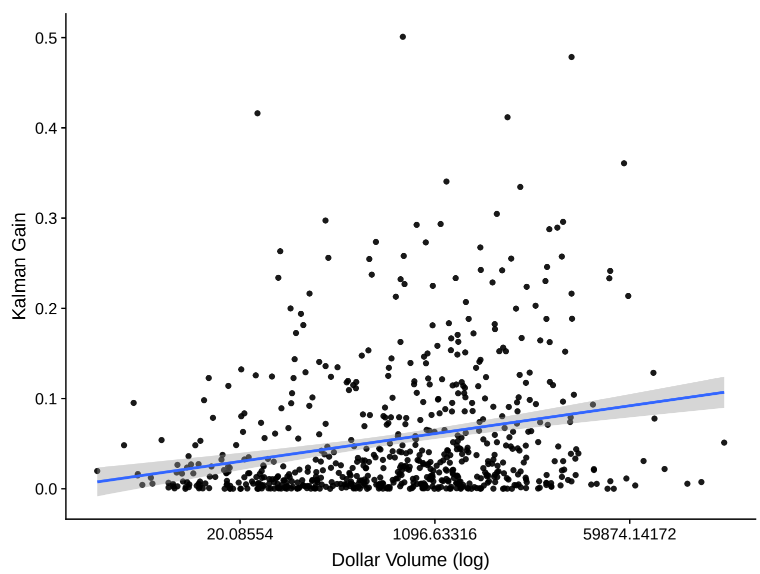

Figure 6: Distribution of Kalman gains across average traded volume in dollars

Note: Figure 6 shows the histogram of Kalman gains for the final sample of 666 stocks sorted by their average traded volume in dollars. The estimates are computed using quarterly data. The blue solid line represents the linear regression line.

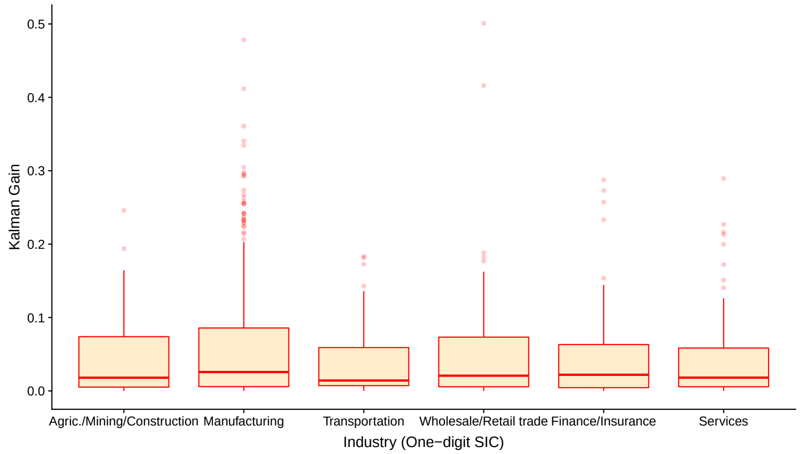

Figure 7: Distribution of Kalman gains by industry

Note: Figure 7 shows the distributions of Kalman gains for the final sample of 666 stocks sorted by industry using one-digit standard industry codes. The estimates are computed using quarterly data.

5.3.4 Informativeness across industries

Figure 7 shows the distribution of Kalman gains in different industries. Our results show heterogeneity in price informativeness across industries and within industry. We find that on average, price informativeness is highest in the manufacturing and agriculture, mining and construction sectors, and lowest in the finance and insurance and services sectors. As Figure 7 illustrates, there is considerable dispersion in price informativeness within each sector. Without a specific conjecture of why one sector would have higher informativeness than other, the main conclusion of this analysis is that there is not a strong relation between industry and price informativeness.

6 Conclusion

We have shown that the outcomes of linear regressions of prices on fundamentals are sufficient to recover exact measures of the ability of asset prices to aggregate dispersed information under minimal assumptions about the environment and exclusively using aggregate market information. In our empirical exercise, we find that the amount of information that can be found by an external observer by conditioning on the price is on average small, although there is a substantial cross sectional dispersion. We find that price informativeness is higher for stocks traded in the NYSE, with higher market capitalization, and traded more frequently. Given that our empirical implementation delivers stock-specific measures of informativeness, there is scope to further explore the relation between informativeness and other outcomes in the cross-section of stocks.

Looking forward, identifying model primitives when there is feedback between financial markets and investment is a challenging but fascinating area for future research, since the nature of production introduces unavoidable non-linearities when there is two-way feedback between financial and real markets.

References

Bai, Jennie, Thomas Philippon, and Alexi Savov. 2015. "Have financial markets become more informative?" *Journal of Financial Economics, Forthcoming.*

Barberis, Nicholas C. 2018. "Psychology-based Models of Asset Prices and Trading Volume."

Bond, Philip, Alex Edmans, and Itay Goldstein. 2012. "The Real Effects of Financial Markets." *Annual Review of Financial Economics*, 4(1): 339–360.

Campbell, John Y. 2017. *Financial decisions and markets: a course in asset pricing.* Princeton University Press.

Chen, Qi, Itay Goldstein, and Wei Jiang. 2006. "Price informativeness and investment sensitivity to stock price." *The Review of Financial Studies*, 20(3): 619–650.

Davila, Eduardo, and Cecilia Parlatore. 2017. "Trading Costs and Informational Efficiency." *Working Paper.*

Davila, Eduardo, and Cecilia Parlatore. 2018. "Volatility and Informativeness." *Working Paper.*

De Long, J Bradford, Andrei Shleifer, Lawrence H Summers, and Robert J Waldmann. 1990. "Noise Trader Risk in Financial Markets." *Journal of Political Economy*, 98(4): 703–738.

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

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

Easley, D., M. O'Hara, and J. Paperman. 1998. "Financial analysts and information-based trade." *Journal of Financial Markets*, 1: 175–201.

Farboodi, Maryam, Adrien Matray, and Laura Veldkamp. 2017. "Where Has All the Big Data Gone?" *Working Paper.*

Gennaioli, Nicola, and Andrei Shleifer. 2018. *A Crisis of Beliefs: Investor Psychology and Financial Fragility.* Princeton University Press.

Granger, Clive WJ, and Paul Newbold. 1974. "Spurious regressions in econometrics." *Journal of Econometrics*, 2(2): 111–120.

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

Hayek, FA. 1945. "The Use of Knowledge in Society." *The American Economic Review*, 35(4): 519–530.

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

Hou, Kewei, Lin Peng, and Wei Xiong. 2013. "Is R2 a Measure of Market Inefficiency?" *working paper.*

Kacperczyk, Marcin, Savitar Sundaresan, and Tianyu Wang. 2018. "Do Foreign Investors Improve Market Efficiency?" *working paper.*

Morck, Randall, Bernard Yeung, and Wayne Yu. 2000. “Tehe information content of stock markets: Why do emerging markets have synchronous stock price movements?” *Journal of Financial Economics*, 58: 215–260.

Roll, Richard. 1988. “R2.” *The Journal of Finance*, 43(3): 541–566.

Veldkamp, Laura. 2009. “Information Choice in Macroeconomics and Finance.” *Manuscript, New York University.*

Vives, Xavier. 2008. *Information and Learning in Markets: the Impact of Market Microstructure.* Princeton University Press.

Vuolteenaho, Tuomo. 2002. “What Drives Firm-Level Stock Returns?” *The Journal of Finance*, 57(1): 233–264.

Weller, Brian M. 2018. “Does Algorithmic Trading Reduce Information Acquisition?” *The Review of Financial Studies*, 31(6): 2184–2226.

Wurgler, Jeffrey. 2000. “Financial markets and the allocation of capital.” *Journal of Financial Economics*, 58: 187–214.

Appendix

A Proofs: Section 2

Assumptions 2 (linear demands) and market clearing imply that

$$\int \Delta q_t^i di = \theta_t \int \alpha_\theta^i di + \int \alpha_s^i s_t^i di + \int \alpha_n^i n_t^i di - p_t \int \alpha_p^i di + \int \psi^i di = 0,$$

so the equilibrium price must satisfy

$$p_t = \frac{\int \alpha_s^i s_t^i di}{\overline{\alpha_p}} + \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t + \frac{\int \alpha_n^i n_t^i di}{\overline{\alpha_p}} + \frac{\overline{\psi}}{\overline{\alpha_p}},$$

where we define cross sectional averages $\overline{\alpha_\theta} = \int \alpha_\theta^i di$, $\overline{\alpha_p} = \int \alpha_p^i di$, and $\overline{\psi} = \int \psi^i di$. Under Assumption 1, which implies the additive structure of signals in Equations. (1) and (2), we can further write

$$p_t = \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\eta_t + \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t + \frac{\overline{\alpha_n}}{\overline{\alpha_p}}n_t + \frac{\int \alpha_s^i \varepsilon_{st}^i di}{\overline{\alpha_p}} + \frac{\int \alpha_n^i \varepsilon_{nt}^i di}{\overline{\alpha_p}} + \frac{\overline{\psi}}{\overline{\alpha_p}},$$

where we define cross sectional averages $\overline{\alpha_s} = \int \alpha_s^i di$ and $\overline{\alpha_n} = \int \alpha_n^i di$. Under a Strong Law of Large Numbers, see, e.g. Vives (2008), the terms $\frac{\int \alpha_s^i \varepsilon_{st}^i di}{\overline{\alpha_p}}$ and $\frac{\int \alpha_n^i \varepsilon_{nt}^i di}{\overline{\alpha_p}}$ vanish when there is a continuum of investors, which allows us to derive Equation (3) in the text.

Proof of Proposition 1. (Identifying absolute price informativeness)

Comparing Equations (7) and (R1) allows us to structurally interpret the coefficients β_0 and β_2 and the residual term as $\beta_0 = \frac{\overline{\psi}}{\overline{\alpha_p}}$, $\beta_2 = \frac{\overline{\alpha_s}}{\overline{\alpha_p}}$, and $\varepsilon_t = \frac{\overline{\alpha_n}}{\overline{\alpha_p}} n_t$. Consequently, $\mathbb{V}\mathrm{ar}\left[\varepsilon_t\right] = \left(\frac{\overline{\alpha_n}}{\overline{\alpha_p}}\right)^2 \tau_n^{-1}$. It then follows that

$$\tau_{\hat{p}} = \frac{\beta_2^2}{\sigma_\varepsilon^2} = \frac{\left(\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\right)^2}{\left(\frac{\overline{\alpha_n}}{\overline{\alpha_p}}\right)^2 \tau_n^{-1}} = \left(\frac{\overline{\alpha_s}}{\overline{\alpha_n}}\right)^2 \tau_n.$$

Proof of Proposition 2. (Identifying relative price informativeness)

We reproduce here Regressions R1 and R2:

$$p_t = \beta_0 + \beta_1 \theta_t + \beta_2 \theta_{t+1} + \varepsilon_t \quad \text{(R1)}$$

$$p_t = \zeta_0 + \zeta_1 \theta_t + \varepsilon_t^\zeta. \quad \text{(R2)}$$

Note that Regression R1 can also be written as

$$p_t = \beta_0 + \left(\beta_1 + \rho\beta_2\right)\theta_t + \beta_2 \eta_t + \varepsilon_t,$$

which allows us to interpret the coefficient ζ_1 in Regression R2 as follows

$$\zeta_1 = \beta_1 + \rho\beta_2,$$

since θ_t is orthogonal to $\varepsilon_t^\zeta = \beta_2 \eta_t + \varepsilon_t$.

Consequently, we can find the following variance decomposition

$$\mathbb{V}\text{ar}\left(p_t\right) = \underbrace{\left(\beta_1 + \rho\beta_2\right)^2}_{\zeta_1^2} \mathbb{V}\text{ar}\left(\theta_t\right) + \beta_2^2 \mathbb{V}\text{ar}\left(\eta_t\right) + \mathbb{V}\text{ar}\left(\varepsilon_t\right),$$

which implies that

$$1 = \frac{\zeta_1^2 \mathbb{V}\text{ar}\left(\theta_t\right)}{\mathbb{V}\text{ar}\left(p_t\right)} + \frac{\beta_2^2 \mathbb{V}\text{ar}\left(\eta_t\right)}{\mathbb{V}\text{ar}\left(p_t\right)} + \frac{\mathbb{V}\text{ar}\left(\varepsilon_t\right)}{\mathbb{V}\text{ar}\left(p_t\right)} \iff 1 = \underbrace{\frac{\zeta_1^2 \mathbb{V}\text{ar}\left(\theta_t\right)}{\mathbb{V}\text{ar}\left(p_t\right)}}_{R^2_{|\theta_t}} + \underbrace{\frac{\mathbb{V}\text{ar}\left(\varepsilon_t\right)}{\mathbb{V}\text{ar}\left(p_t\right)}}_{1-R^2_{|\theta_{t+1},\theta_t}} \left(1 + \underbrace{\frac{\beta_1^2 \mathbb{V}\text{ar}\left(\eta_t\right)}{\mathbb{V}\text{ar}\left(\varepsilon_t\right)}}_{\tau^R_{\hat{p}}}\right).$$

Therefore,

$$1 - R^2_{|\theta_t} = \left(1 - R^2_{|\theta_{t+1},\theta_t}\right)\left(1 + \tau^R_{\hat{p}}\right) \Rightarrow \frac{1 - R^2_{|\theta_t}}{1 - R^2_{|\theta_{t+1},\theta_t}} = 1 + \tau^R_{\hat{p}},$$

which implies that

$$\tau^R_{\hat{p}} = \frac{R^2_{|\theta_{t+1},\theta_t} - R^2_{|\theta_t}}{1 - R^2_{|\theta_{t+1},\theta_t}}.$$

Moreover, the Kalman gain can be recovered by

$$K \equiv \frac{\tau_{\hat{p}}}{\tau_{\hat{p}} + \tau_\eta} = \frac{\tau^R_{\hat{p}}}{1 + \tau^R_{\hat{p}}} = \frac{R^2_{|\theta_{t+1},\theta_t} - R^2_{|\theta_t}}{1 - R^2_{|\theta_t}}.$$

Correlated noise and fundamental

In the baseline model we have assumed that the innovation to the asset payoff η_t is uncorrelated with the aggregate source of noise n_t. In this section, we allow for the aggregate source of noise to be correlated with the fundamental. Formally, we consider the following signal structure for the aggregate noise

$$n_t = \omega\eta_t + \varepsilon_{nt},$$

where $\varepsilon_{nt} \sim N\left(\mu_n, \tau_n^{-1}\right)$. Then, under Assumptions 1 and 2, market clearing in the risky asset market implies

$$\begin{aligned} p_t &= \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t + \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\eta_t + \frac{\overline{\alpha_n}}{\overline{\alpha_p}}n_t + \frac{\overline{\psi}}{\overline{\alpha_p}} \\ &= \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t + \frac{\overline{\alpha_s} + \omega\overline{\alpha_n}}{\overline{\alpha_p}}\eta_t + \frac{\overline{\alpha_n}}{\overline{\alpha_p}}\varepsilon_{nt} + \frac{\overline{\psi}}{\overline{\alpha_p}}. \end{aligned} \tag{29}$$

From the perspective of an external observer who only sees the price and the current payoff of the asset, the unbiased signal of the fundamental contained in the price is given by

$$\hat{p} = \frac{\overline{\alpha_p}}{\overline{\alpha_s}}\left(p - \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t - \frac{\overline{\alpha_n}}{\overline{\alpha_p}}\mu_n - \frac{\overline{\psi}}{\overline{\alpha_p}}\right)$$

and absolute price informativeness is given by

$$\tau_{\hat{p}} \equiv \left(\mathbb{V}\text{ar}\left[\hat{p}_t | \theta_{t+1}, \theta_t\right]\right)^{-1} = \left(\frac{\overline{\alpha_s} + \omega\overline{\alpha_n}}{\overline{\alpha_n}}\right)^2 \tau_n.$$

Note that mapping the coefficients in Regression R1 to the equilibrium price in Eq. 29 implies that $\beta_2 = \frac{\overline{\alpha_s}+\omega\overline{\alpha_n}}{\overline{\alpha_p}}$ and $\varepsilon_t = \frac{\overline{\alpha_n}}{\overline{\alpha_p}}\varepsilon_{nt}$. Then, we can recover absolute and relative price informativeness in the same way as in the baseline model as follows

$$\tau_{\hat{p}} = \frac{\beta_2^2}{\sigma_\varepsilon^2} \quad \text{and} \quad \tau_{\hat{p}}^R = \frac{R^2_{|\theta_{t+1},\theta_t} - R^2_{|\theta_t}}{1 - R^2_{|\theta_{t+1},\theta_t}},$$

where the steps to recover relative price informativeness are exactly the same as those in the baseline model. Therefore, we can assume that the private trading motives are orthogonal to the asset payoff without loss of generality.

B Proofs: Section 3

Equilibrium Characterization

An investor i born in period t chooses a quantity of the risky asset to solve

$$\max_{q_t^i} \left(\mathbb{E}\left[\theta_{t+1} + R^{-1}p_{t+1}|\mathcal{I}_t^i\right] - p_t\right) q_t^i - \frac{\gamma}{2}\mathbb{V}ar\left[\theta_{t+1} + R^{-1}p_{t+1}|\mathcal{I}_t^t\right]\left(q_t^i\right)^2,$$

where $\mathcal{I}_t^i = \{\theta_t, s_t^i, \overline{\eta}_t^i, p_t\}$ is the information set of investor i at time t.

The first order condition for an investor at time t is

$$q_t^i = \frac{\mathbb{E}\left[\theta_{t+1} + R^{-1}p_{t+1}|\mathcal{I}_t^i\right] - p_t}{\gamma\mathbb{V}ar\left[\theta_{t+1} + R^{-1}p_{t+1}|\mathcal{I}_t^i\right]}.$$

In an equilibrium in linear strategies, we conjecture and verify that the demand of an investor i can be written as

$$\Delta q_t^i = \alpha_\theta^i\theta_t + \alpha_s^i s_{it} + \alpha_\eta^i\overline{\eta}_t^i - \alpha_p^i p_t + \psi^i.$$

Market clearing and the Strong Law of Large Numbers imply

$$p_t = \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t + \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\eta_t + \frac{\overline{\alpha_\eta}}{\overline{\alpha_p}}n_t + \frac{\overline{\psi}}{\overline{\alpha_p}}.$$

The unbiased signal of the innovation in the dividend contained in the price is

$$\hat{p}_t = \frac{\overline{\alpha_p}}{\overline{\alpha_s}}\left(p_t - \frac{\overline{\alpha_\eta}}{\overline{\alpha_s}}\mu_n - \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t - \frac{\overline{\psi}}{\overline{\alpha_p}}\right) = \eta_t + \frac{\overline{\alpha_\eta}}{\overline{\alpha_s}}\left(n_t - \mu_n\right),$$

where

$$\hat{p}_t|\theta_{t+1},\theta_t \sim N\left(\eta_t, \tau_{\hat{p}}^{-1}\right),$$

with price informativeness given by

$$\tau_{\hat{p}} = \left(\mathbb{V}\text{ar}\left[\hat{p}_t|\theta_{t+1},\theta_t\right]\right)^{-1} = \left(\frac{\overline{\alpha_s}}{\overline{\alpha_\eta}}\right)^2\tau_{\overline{\eta}}.$$

Given our guesses for the demand functions and the linear structure of prices we have

$$\theta_{t+1} + R^{-1}p_{t+1} = \theta_{t+1} + R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_{t+1} + R^{-1}\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\eta_{t+1} + R^{-1}\frac{\overline{\alpha_\eta}}{\overline{\alpha_p}}n_{t+1} + R^{-1}\frac{\overline{\psi}}{\overline{\alpha_p}},$$

$$\mathbb{E}\left[\theta_{t+1}+R^{-1}p_{t+1}|\mathcal{I}_t^i\right]=\left(1+R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)\mathbb{E}\left[\theta_{t+1}|\mathcal{I}_t^i\right]+R^{-1}\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\mathbb{E}\left[\eta_{t+1}\right]+R^{-1}\frac{\overline{\alpha_\eta}}{\overline{\alpha_p}}\mathbb{E}\left[n_{t+1}\right]+R^{-1}\frac{\overline{\psi}}{\overline{\alpha_p}}$$
$$=\left(1+R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)\left(\rho\theta_t+\mathbb{E}\left[\eta_t|\mathcal{I}_t^i\right]\right)+R^{-1}\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\mathbb{E}\left[\eta_{t+1}\right]+R^{-1}\frac{\overline{\alpha_\eta}}{\overline{\alpha_p}}\overline{n_t}+R^{-1}\frac{\overline{\psi}}{\overline{\alpha_p}},$$

and

$$\mathbb{V}ar\left[\theta_{t+1}+R^{-1}p_{t+1}|\mathcal{I}_t^i\right]=\left(1+R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)^2\mathbb{V}ar\left[\theta_{t+1}|\mathcal{I}_t^i\right]+\left(R^{-1}\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\right)^2\mathbb{V}ar\left[\eta_{t+1}\right]+\left(R^{-1}\frac{\overline{\alpha_\eta}}{\overline{\alpha_p}}\right)^2\mathbb{V}ar\left[n_{t+1}\right]$$
$$=\left(1+R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)^2\mathbb{V}ar\left[\eta_t|\mathcal{I}_t^i\right]+\left(R^{-1}\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\right)^2\mathbb{V}ar\left[\eta_{t+1}\right]+\left(R^{-1}\frac{\overline{\alpha_\eta}}{\overline{\alpha_p}}\right)^2\tau_{nt+1}^{-1}.$$

Moreover, given the Gaussian structure of the signals in the information set, Bayesian updating implies

$$\mathbb{E}\left[\eta_t|s_{t^i},\overline{\eta}_t^i,p_t\right]=\frac{\tau_s s_t^i+\tau_\eta\overline{\eta}_t^i+\tau_{\hat{p}}\hat{p}_t}{\tau_s+\tau_\eta+\tau_{\hat{p}}}=\frac{\tau_s s_t^i+\tau_\eta\overline{\eta}_t^i++\tau_{\hat{p}}\frac{\overline{\alpha_p}}{\overline{\alpha_s}}\left(p_t-\frac{\overline{\alpha_\eta}}{\overline{\alpha_s}}\mu_n-\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t-\frac{\overline{\psi}}{\overline{\alpha_p}}\right)}{\tau_s+\tau_\eta+\tau_{\hat{p}}},$$

and

$$\mathbb{V}ar\left[\eta_t|\mathcal{I}_t^i\right]=\mathbb{V}ar\left[\eta_t|s_t^i,\overline{\eta}_t^i,p_t\right]=\left(\tau_s+\tau_\eta+\tau_{\hat{p}}\right)^{-1}.$$

Then, the first order condition is the given by is

$$q_t^i=\frac{1}{\gamma}\frac{\left(1+R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)\left(\rho\theta_t+\mathbb{V}ar\left[\eta_t|\mathcal{I}_t^i\right]\left(\tau_s s_t^i+\tau_\eta\overline{\eta}_t^i+\tau_{\hat{p}}\frac{\overline{\alpha_p}}{\overline{\alpha_s}}\left(p_t-\frac{\overline{\alpha_\eta}}{\overline{\alpha_s}}\mu_n-\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t-\frac{\overline{\psi}}{\overline{\alpha_p}}\right)\right)\right)+R^{-1}\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\mathbb{E}\left[\eta_{t+1}\right]+R^{-1}\frac{\overline{\alpha_\eta}}{\overline{\alpha_p}}\mu_n+R^{-1}\frac{\overline{\psi}}{\overline{\alpha_p}}-p_t}{\left(1+R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)^2\mathbb{V}ar\left[\eta_t|\mathcal{I}_{it}\right]+\left(R^{-1}\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\right)^2\mathbb{V}ar\left[\eta_{t+1}\right]+\left(R^{-1}\frac{\overline{\alpha_\eta}}{\overline{\alpha_p}}\right)^2\tau_n^{-1}}.$$

Matching coefficients we have

$$\alpha_s^i=\frac{\left(1+R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)}{\kappa}\mathbb{V}ar\left[\eta_t|\mathcal{I}_t^i\right]\tau_s \tag{30}$$
$$\alpha_\eta^i=\frac{\left(1+R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)}{\kappa}\mathbb{V}ar\left[\eta_t|\mathcal{I}_t^i\right]\tau_\eta$$
$$\alpha_\theta^i=\frac{\left(1+R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)}{\kappa}\left(\rho-\mathbb{V}ar\left[\eta_t|\mathcal{I}_t^i\right]\tau_{\hat{p}}\frac{\overline{\alpha_\theta}}{\overline{\alpha_s}}\right)$$
$$\alpha_p^i=\frac{1}{\kappa}\left(1-\left(1+R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)\mathbb{V}ar\left[\eta_t|\mathcal{I}_t^i\right]\tau_{\hat{p}}\frac{\overline{\alpha_p}}{\overline{\alpha_s}}\right)$$
$$\psi^i=-\frac{1}{\kappa}\left(\left(1+R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)\mathbb{V}ar\left[\eta_t|\mathcal{I}_t^i\right]\tau_{\hat{p}}\left(\frac{\overline{\alpha_\eta}}{\overline{\alpha_s}}\mu_n+\frac{\overline{\psi}}{\overline{\alpha_s}}\right)-R^{-1}\left(\frac{\overline{\alpha_\eta}}{\overline{\alpha_p}}\mu_n+\frac{\overline{\psi}}{\overline{\alpha_p}}\right)\right),$$

where

$$\kappa\equiv\gamma\left(\left(1+R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)^2\mathbb{V}ar\left[\eta_t|\mathcal{I}_t^i\right]+\left(R^{-1}\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\right)^2\mathbb{V}ar\left[\eta_{t+1}\right]+\left(R^{-1}\frac{\overline{\alpha_\eta}}{\overline{\alpha_p}}\right)^2\tau_n^{-1}\right),$$

since $\mathbb{V}ar\left[\eta_t|\mathcal{I}_t^i\right]=\left(\tau_s+\tau_\eta+\tau_{\hat{p}}\right)^{-1}$ for all i.

Then, an equilibrium in linear strategies always exists if the system above has a solution. Note that the demand sensitivities are the same for all i. Then, there exists a unique solution to the system in Equations (30), that is given by

$$\alpha_s^i=\frac{1}{\kappa}\frac{1}{1-R^{-1}\rho}\frac{\tau_s}{\tau_\eta+\tau_s+\tau_{\hat{p}}},\quad \alpha_\eta^i=\frac{1}{\kappa}\frac{1}{1-R^{-1}\rho}\frac{\tau_\eta}{\tau_\eta+\tau_s+\tau_{\hat{p}}}$$
$$\alpha_\theta^i=\frac{1}{\kappa}\frac{\rho}{1-R^{-1}\rho}\frac{\tau_s}{\tau_s+\tau_{\hat{p}}},\quad \alpha_p^i=\frac{1}{\kappa}\frac{\tau_s}{\tau_s+\tau_{\hat{p}}},\quad\text{and}$$
$$\psi^i=-\frac{\frac{1}{\kappa}\frac{1}{1-R^{-1}\rho}\left(\left(1-R^{-1}\right)\tau_{\hat{p}}-R^{-1}\tau_s\right)\frac{\tau_\eta}{\tau_\eta+\tau_s+\tau_{\hat{p}}}\mu_n}{\left(1+\left(1-R^{-1}\right)\tau_{\hat{p}}-R^{-1}\tau_s\right)},$$

where

$$\tau_{\hat{p}} = \left(\frac{\tau_s}{\tau_\eta}\right)^2 \tau_{\overline{\eta}}$$

and

$$\kappa = \gamma\left(\left(\frac{1}{1-R^{-1}\rho}\right)^2 \frac{1}{\tau_\eta+\tau_s+\tau_{\hat{p}}} + \left(R^{-1}\frac{1}{1-R^{-1}\rho}\frac{\tau_s+\tau_{\hat{p}}}{\tau_\eta+\tau_s+\tau_{\hat{p}}}\right)^2 \tau_\eta^{-1} + \left(\frac{R^{-1}}{1-R^{-1}\rho}\frac{\tau_s+\tau_{\hat{p}}}{\tau_\eta+\tau_s+\tau_{\hat{p}}}\frac{\tau_\eta}{\tau_s}\right)^2 \tau_{\overline{\eta}}^{-1}\right).$$

Proof of Proposition 3. (Identifying signal and trading needs precisions)

We reproduce here Regressions R1 and R2:

$$p_t = \beta_0 + \beta_1\theta_t + \beta_2\theta_{t+1} + \varepsilon_t \tag{R1}$$

$$p_t = \zeta_0 + \zeta_1\theta_t + \varepsilon_t^{\zeta}. \tag{R2}$$

Note that Regression R1 can be written as

$$p_t = \beta_0 + (\beta_1 + \rho\beta_2)\,\theta_t + \beta_2\eta_t + \varepsilon_t,$$

and that the following equilibrium equation is valid in this model

$$p_t = \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t + \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\eta_t + \frac{\overline{\alpha_\eta}}{\overline{\alpha_p}}n_t + \frac{\overline{\psi}}{\overline{\alpha_p}}.$$

From the characterization of the equilibrium described above, we can express $\frac{\alpha_s}{\alpha_p}$ and $\frac{\alpha_s}{\alpha_\eta}$ as follows:

$$\frac{\overline{\alpha_s}}{\overline{\alpha_p}} = \left(1 + R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)\frac{\tau_s+\tau_{\hat{p}}}{\tau_\eta+\tau_s+\tau_{\hat{p}}} = \left(1 + R^{-1}\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\right)\frac{\frac{\tau_s}{\tau_\eta}+\tau_{\hat{p}}^R}{1+\frac{\tau_s}{\tau_\eta}+\tau_{\hat{p}}^R} \tag{31}$$

$$\frac{\overline{\alpha_s}}{\overline{\alpha_\eta}} = \frac{\tau_s}{\tau_\eta}.$$

Under the stated assumptions, we can therefore interpret the coefficients β_1 and $\zeta_1$as follows

$$\beta_2 = \frac{\overline{\alpha_s}}{\overline{\alpha_p}} \quad \text{and} \quad \zeta_1 = \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}. \tag{32}$$

Therefore, Equations (31) and Equation (32) imply that $\frac{\tau_s}{\tau_\eta} + \tau_{\hat{p}}^R$ can be recovered as follows

$$\beta_2 = \left(1 + R^{-1}\zeta_1\right)\frac{\frac{\tau_s}{\tau_\eta}+\tau_{\hat{p}}^R}{1+\frac{\tau_s}{\tau_\eta}+\tau_{\hat{p}}^R} \Rightarrow \frac{\tau_s}{\tau_\eta} + \tau_{\hat{p}}^R = \frac{\beta_2}{1+R^{-1}\zeta_1-\beta_2},$$

which allows to express τ_s:

$$\tau_s = \frac{\beta_2}{1+R^{-1}\zeta_1-\beta_2}\tau_\eta - \tau_{\hat{p}} = \tau_{\hat{p}}\left(\frac{\beta_2}{1+R^{-1}\zeta_1-\beta_2}\frac{1}{\tau_{\hat{p}}^R} - 1\right),$$

using the fact that τ_η can be recovered from $\tau_\eta = \frac{\tau_{\hat{p}}}{\tau_{\hat{p}}^R}$. Finally, exploiting the relation $\tau_{\hat{p}} = \left(\frac{\tau_s}{\tau_\eta}\right)^2 \tau_n$, τ_n can be recovered as follows

$$\tau_{\hat{p}} = \left(\frac{\tau_s}{\tau_\eta}\right)^2 \tau_n = \left(\frac{\beta_2}{1+R^{-1}\zeta_1-\beta_2} - \tau_{\hat{p}}^R\right)^2 \tau_n \Rightarrow \tau_n = \left(\frac{\beta_2}{1+R^{-1}\zeta_1-\beta_2} - \tau_{\hat{p}}^R\right)^{-2} \tau_{\hat{p}},$$

where R can be mapped to the risk-free rate.

C Proofs: Section 4

Proof of Proposition 4. (Identifying absolute price informativeness: random walk)

We reproduce here Regression R4 and the equilibrium condition in Equation (11):

$$\Delta p_t = \beta_1 \Delta\theta_t + \beta_2 \Delta\theta_{t+1} + \varepsilon_t, \tag{R4}$$

$$\Delta p_t = \left(\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}} - \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\right)\Delta\theta_t + \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\Delta\theta_{t+1} + \frac{\overline{\alpha_n}}{\overline{\alpha_p}}\Delta n_t. \tag{11}$$

Consequently, we can express $\mathbb{V}\mathrm{ar}\left[\varepsilon_t\right]$ as follows

$$\mathbb{V}ar\left[\varepsilon_t\right] = \left(\frac{\overline{\alpha_n}}{\overline{\alpha_p}}\right)^2 \mathbb{V}ar\left[\Delta n_t\right] = \left(\frac{\overline{\alpha_n}}{\overline{\alpha_p}}\right)^2 2\tau_n^{-1},$$

where the last equality follows from the i.i.d. assumption of n_t shocks. We can therefore find $\tau_{\hat{p}}$ as

$$\tau_{\hat{p}} = 2\frac{\beta_2^2}{\sigma_\varepsilon^2} = \frac{2\left(\frac{\overline{\alpha_s}}{\overline{\alpha_p}}\right)^2}{\left(\frac{\overline{\alpha_n}}{\overline{\alpha_p}}\right)^2 2\tau_n^{-1}} = \left(\frac{\overline{\alpha_s}}{\overline{\alpha_n}}\right)^2 \tau_n.$$

Proof of Proposition 5. (Identifying relative price informativeness: random walk)

In this model, the equilibrium pricing equation can be written as

$$\begin{aligned} p_t &= \frac{\overline{\alpha_\theta}}{\overline{\alpha_p}}\theta_t + \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\eta_t + \frac{\overline{\alpha_n}}{\overline{\alpha_p}}n_t + \frac{\overline{\psi}}{\overline{\alpha_p}} \\ &= \left(\frac{\overline{\alpha_\theta}}{\overline{\alpha_p}} - \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\right)\theta_t + \frac{\overline{\alpha_s}}{\overline{\alpha_p}}\theta_{t+1} + \frac{\overline{\alpha_n}}{\overline{\alpha_p}}n_t + \frac{\overline{\psi}}{\overline{\alpha_p}} + \mu_\theta. \end{aligned}$$

Taking first differences, this Equation yields Equation (11) in the text, which yields consistent estimates when estimated by OLS, since $\Delta\theta_t = \eta_{t-1}$ and $\Delta\theta_{t+1} = \eta_t$ are orthogonal to the error term. We exploit Regressions R4 and R5, reproduced here.

$$\Delta p_t = \beta_1 \Delta\theta_t + \beta_2 \Delta\theta_{t+1} + \varepsilon_t, \tag{R4}$$

$$\Delta p_t = \zeta_0 + \zeta_1 \Delta\theta_t + \varepsilon_t^\zeta. \tag{R5}$$

Note that $\zeta_1 = \beta_1$, given that η_{t-1} is orthogonal to ε_t^ζ.

Exploiting Regression R4, we can write the following variance decomposition

$$\mathbb{V}\mathrm{ar}\left[\Delta p_t\right] = \beta_1^2 \mathbb{V}\mathrm{ar}\left[\Delta\theta_t\right] + \beta_2^2 \mathbb{V}\mathrm{ar}\left[\Delta\theta_{t+1}\right] + \mathbb{V}\mathrm{ar}\left[\varepsilon_t\right],$$

which when divided by $\mathbb{V}\mathrm{ar}\left[\Delta p_t\right]$ yields

$$1 = \underbrace{\frac{\beta_1^2 \mathbb{V}\mathrm{ar}\left[\Delta\theta_t\right]}{\mathbb{V}\mathrm{ar}\left[\Delta p_t\right]}}_{R^2_{|\Delta\theta_t}} + \underbrace{\frac{\mathbb{V}\mathrm{ar}\left[\varepsilon_t\right]}{\mathbb{V}\mathrm{ar}\left[\Delta p_t\right]}}_{1-R^2_{|\Delta\theta_{t+1},\Delta\theta_t}} \left(\underbrace{\frac{\beta_2^2 \mathbb{V}\mathrm{ar}\left[\Delta\theta_{t+1}\right]}{\mathbb{V}\mathrm{ar}\left[\varepsilon_t\right]}}_{\frac{1}{2}\tau_{\hat{p}}^R} + 1\right).$$

This expression can be rearranged to solve for $\tau_{\hat{p}}^R$ as follows:

$$\frac{1-R^2_{|\Delta\theta_t}}{1-R^2_{|\Delta\theta_{t+1},\Delta\theta_t}} = 1+\frac{1}{2}\tau_{\hat{p}}^R \Rightarrow \tau_{\hat{p}}^R = 2\frac{R^2_{|\Delta\theta_{t+1},\Delta\theta_t}-R^2_{|\Delta\theta_t}}{1-R^2_{|\Delta\theta_{t+1},\Delta\theta_t}},$$

which corresponds to Equation (13) in the text. The Kalman Gain for an external observer can be calculated as $K=\frac{\tau_{\hat{p}}^R}{1+\tau_{\hat{p}}^R}$.

Multiple Risky Assets

The price of asset j is given by

$$\begin{aligned}
p_{j,t} &= \left[\left(\overline{A_p}\right)^{-1}\left(\overline{A_\theta}-\overline{A_s}C\right)\right]_j \boldsymbol{\theta}_t + \left[\left(\overline{A_p}\right)^{-1}\overline{A_s}\right]_j \boldsymbol{\theta}_{t+1} + \left[\left(\overline{A_p}\right)^{-1}\overline{A_n}\right]_j \boldsymbol{n}_t + \left[\left(\overline{A_p}\right)^{-1}\overline{A_0}\right]_j \\
&= \sum_{h=1}^{N}\left(\left[\left(\overline{A_p}\right)^{-1}\left(\overline{A_\theta}-\overline{A_s}C\right)\right]_{jh}\theta_{h,t} + \left[\left(\overline{A_p}\right)^{-1}\overline{A_s}\right]_{jh}\theta_{h,t+1} + \left[\left(\overline{A_p}\right)^{-1}\overline{A_n}\right]_{jh} n_{h,t}\right) + \left[\left(\overline{A_p}\right)^{-1}\overline{A_0}\right]_j \\
&= \sum_{h}\pi_{jh}\left(\left[\left(\overline{A_p}\right)^{-1}\left(\overline{A_\theta}-\overline{A_s}C\right)\right]_{jh}\theta_{j,t} + \left[\left(\overline{A_p}\right)^{-1}\overline{A_s}\right]_{jh}\theta_{j,t+1}\right) + \sum_{h\neq j}\left(\left[\left(\overline{A_p}\right)^{-1}\left(\overline{A_\theta}-\overline{A_s}C\right)\right]_{jh}\omega^j_{h,t-1} + \left[\left(\overline{A_p}\right)^{-1}\overline{A_s}\right]_{jh}\omega_{h,t}\right) \\
&\quad + \sum_{h}\left[\left(\overline{A_p}\right)^{-1}\overline{A_n}\right]_{jh} n_{h,t} + \left[\left(\overline{A_p}\right)^{-1}\overline{A_0}\right]_j,
\end{aligned}$$

where $\theta_{h,t+1}=\pi_{jh}\theta_{j,t+1}+\omega^j_{h,t}$ with $\pi_{jj}=1$ and $\omega^j_{j,t}=0$. Hence, the estimates in the regression

$$p_{j,t}=\beta_0+\beta_1\theta_{j,t}+\beta_2\theta_{j,t+1}+\varepsilon_{j,t}$$

imply

$$\begin{aligned}
\varepsilon_{jt} &= \sum_{h\neq j}\left(\left[\left(\overline{A_p}\right)^{-1}\left(\overline{A_\theta}-\overline{A_s}C\right)\right]_{jh}\omega^j_{h,t-1} + \left[\left(\overline{A_p}\right)^{-1}\overline{A_s}\right]_{jh}\omega_{h,t}\right) + \sum_{h=1}^{N}\left[\left(\overline{A_p}\right)^{-1}\overline{A_n}\right]_{jh} n_{h,t} \\
\beta_2 &= \sum_{h=1}^{N}\left[\left(\overline{A_p}\right)^{-1}\overline{A_s}\right]_{jh}\pi_{jh}.
\end{aligned}$$

Therefore, price informativeness can be recovered as

$$\hat{\tau}^j_{\hat{p}} = \frac{\beta_2^2}{\mathbb{V}\mathrm{ar}\left[\varepsilon_{j,t}\right]}.$$

D Additional Results

In this Section, we report additional results mentioned in the text. Table A1 reports the distribution of our estimates for absolute price informativeness, relative price informativeness and Kalman gains assuming all stocks have non-stationary earning processes using quarterly data. Table A2 provides summary statistics of our data at an annual frequency. Finally, Table and the corresponding estimates of our measures of price informativeness.

Table A1: Results assuming non-stationary earnings

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Absolute Price Informativeness	666	0.04	0.25	0.00	0.0001	0.001	0.01	4.51
Relative Price Informativeness	666	0.08	0.12	0.00	0.01	0.03	0.10	1.00
Kalman Gain	666	0.06	0.08	0.00	0.01	0.03	0.09	0.50

Note: Table A1 presents summary statistics on the informative measures recovered assuming the earnings process for all stocks is non stationary using quarterly data. It provides information on the sample mean, median, and standard deviation, as well as the 25th and 75th percentiles of the A3 for the sample without outliers.

Table A2: Summary statistics for annual data

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Mean Earnings	466	778.08	2,038.33	−7.04	41.41	170.12	567.11	21,649.80
St. Dev. Earnings	466	662.11	1,771.28	1.73	32.05	136.22	531.08	15,434.57

Table A2 presents summary statistics for the full sample of 466 stocks at an annual frequency. It provides information on the sample mean, median, and standard deviation, as well as the minimum, the maximum and the 25th and 75th percentiles of the distribution of the variance of earnings. All variables are expressed in millions of dollars in 2008.

Table A3: Estimates of Price Informativeness at an Annual Frequency

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Absolute Price Informativeness	272	0.004	0.02	0.00	0.0000	0.0001	0.001	0.29
Relative Price Informativeness	272	0.44	0.49	0.0000	0.10	0.28	0.62	3.07
Kalman Gain	272	0.25	0.19	0.0000	0.09	0.22	0.38	0.75

Note: Table A3 presents summary statistics on the informative measures recovered assuming the earnings process for all stocks is non stationary using annual data. It provides information on the sample mean, median, and standard deviation, as well as the 25th and 75th percentiles of the distribution for the sample without outliers.

Online Appendix

A Relation to Bai, Philippon, Savov (2015)

In this section, we explicitly re-derive the model in Bai, Philippon and Savov (2015) (BPS) as a special case of our general framework. Consistent with our approach, we abstract from investment decisions, and exclusively focus on the role of financial markets aggregating information. Identical remarks apply to the subsequent literature that uses forecasting price efficiency, including (Farboodi, Matray and Veldkamp, 2017) and (Kacperczyk, Sundaresan and Wang, 2018).

There are two major differences between our approach and the one developed in BPS. The first and most important difference concerns how price informativeness is measured. The second difference lies in the assumptions behind the empirical implementation.

As we describe in the text, BPS use forecasting price efficiency (FPE), i.e., the unconditional variance of the expected value of the fundamental conditional on the price, as a proxy for price informativeness. While higher price informativeness will lead to higher FPE, higher FPE may not reflect an increase in price informativeness. More specifically, we show that FPE confounds changes in the volatility of the fundamental with changes in the ability of markets to aggregate dispersed information. Hence, FPE can increase because price informativeness increases or because the fundamental becomes more volatile and harder to predict. Alternatively, absolute price informativeness is the precision of the unbiased signal about the fundamental contained in the price. This precision is a direct measure of the ability of financial markets to aggregate dispersed information and it is independent of the volatility of the fundamental.

Regarding the empirical implementation, the regressions run in BPS and the regressions that we run in this paper to recover price informativeness are different. BPS run cross-sectional regressions of fundamentals on prices and report the time-series evolution of their cross-sectional estimates of FPE. BPS provide estimates of FPE at the market level (or for a subset of the market) over time. In contrast, we use time series regressions to provide firm-specific measures of price informativeness. Moreover, we run regressions of prices on fundamentals, which we show to be more adequate to obtain consistent estimates of price informativeness. The main difference between these two approaches concerns the underlying assumptions regarding the nature of firm-specific primitives and investors' characteristics (e.g., volatility of the fundamental, precision of private signals, or volatility of private trading needs) across time and firms. By running cross-sectional regressions, BPS assume that all firms' fundamentals (among others, the volatility of the fundamental and noise, as well as the precision of investors' private signals) are identical in a given period. This assumption is unlikely to hold in practice and clearly rejected by the data (see Table 2). Our approach assumes instead that firm-specific parameters are time-invariant, but allows for firm specific parameters to vary freely in the cross-section of firms.

Measures of price informativeness

In the remained of this section, our approach to the approach in BPS in more detail. To compare FPE to absolute and relative price informativeness we first describe the environment in BPS using our notation to show how it is nested in our general specification. Then, we show that while FPE is relevant for welfare, it does not disentangle the ability of markets to aggregate information from how easy it is to forecast the fundamental.

Environment There are two periods, $t = 0, 1$. There is one asset with a payoff $\theta \sim N\left(\overline{\theta}, \tau_\theta^{-1}\right)$. There are $i = 1, ..., I$ informed traders who choose their demand q_{1i} to maximize mean variance preferences with imperfect information about θ. The asset payoff θ is not observable. However, investors observe a private signal

$$s = \theta + \varepsilon_s$$

and a public signal

$$\pi = \theta + \varepsilon_\pi,$$

where $\varepsilon_s \sim N\left(0, \tau_s^{-1}\right)$, $\varepsilon_\pi \sim N\left(0, \tau_\pi^{-1}\right)$, and $\varepsilon_s \perp \varepsilon_\pi$. Note that all informed investors observe the same set of signals. There are N noise traders whose total demand is random and given by $n \sim N\left(0, \tau_n^{-1}\right)$.

The informed traders' problem is

$$\max_{q_{1i}} \left(\mathbb{E}\left[\theta|s, \pi\right] - p\right) q_{1i} - \frac{\gamma}{2} \mathbb{V}ar\left[\theta|s, \pi\right] q_{1i}^2 + p q_{0i}$$

which leads to the following demand curve

$$q_{1i} = \frac{\mathbb{E}\left[\theta|s, \pi\right] - p}{\gamma \mathbb{V}ar\left[\theta|s, \pi\right]},$$

where

$$\mathbb{E}\left[\theta|s, \eta\right] = \frac{\tau_\theta \overline{\theta} + \tau_s s + \tau_\pi \pi}{\tau_\theta + \tau_s + \tau_\pi} \quad \text{and} \quad \mathbb{V}ar\left[\theta|s, \eta\right] = \frac{1}{\tau_\theta + \tau_s + \tau_\pi}.$$

Since all informed investors share the same information set, there is no learning from the price.

In an equilibrium in linear strategies net demands for informed traders are given by

$$\Delta q_{1i}^I = \alpha_s^I s + \alpha_\pi^I \pi + \alpha_n^I n - \alpha_p^I p + \psi^I,$$

and for uninformed traders

$$\Delta q_{1i}^U = \alpha_s^U s + \alpha_\pi^U \pi + \alpha_n^U n - \alpha_p^U p + \psi^U.$$

Matching coefficients we have that

$$\begin{aligned}
\alpha_s^I &= \frac{\tau_s}{\gamma}, \quad \alpha_\pi^I = \frac{\tau_\pi}{\gamma}, \quad \alpha_n^I = 0, \\
\alpha_p^I &= \frac{1}{\gamma}\left(\tau_\theta + \tau_s + \tau_\pi\right) \\
\psi^I &= \frac{\tau_\theta}{\gamma} \overline{\theta} - q_{0i},
\end{aligned}$$

and $\alpha_s^U = \alpha_\pi^U = \alpha_p^U = \psi^U = 0$, and $\alpha_n^U = \frac{1}{N}$.

Market clearing implies

$$\sum_{i=1}^{I} \Delta q_{si}^I + n = 0,$$

which is the same as

$$p = \frac{\overline{\alpha_s}}{\overline{\alpha_p}} s + \frac{\overline{\alpha_\pi}}{\overline{\alpha_p}} \pi + \frac{\overline{\psi}}{\overline{\alpha_p}} + \frac{\overline{\alpha_n}}{\overline{\alpha_p}} n,$$

where $\overline{\alpha_s} = I\alpha_s^I + N\alpha_s^N$. $\overline{\alpha_\pi} = I\alpha_\pi^I + N\alpha_\pi^N$, $\overline{\alpha_p} = I\alpha_p^I + N\alpha_p^N$, $\overline{\alpha_n} = I\alpha_n^I + N\alpha_n^N$, and $\overline{\psi} = I\psi^I + N\psi^N$.

Price informativeness and forecasting price efficiency Observing the price is equivalent to observing

$$\hat{p} = \frac{\overline{\alpha_p}}{\overline{\alpha_s} + \overline{\alpha_\pi}} \left(p - \frac{\overline{\psi}}{\overline{\alpha_p}} \right) = \theta + \frac{\overline{\alpha_s}}{\overline{\alpha_s} + \overline{\alpha_\pi}} \varepsilon_s + \frac{\overline{\alpha_\pi}}{\overline{\alpha_s} + \overline{\alpha_\pi}} \varepsilon_\pi + \frac{\overline{\alpha_n}}{\overline{\alpha_s} + \overline{\alpha_\pi}} n,$$

where

$$\hat{p} | \theta \sim N\left(\theta, \tau_{\hat{p}}^{-1}\right)$$

with

$$\tau_{\hat{p}} = \left(\frac{\overline{\alpha_s}}{\overline{\alpha_s} + \overline{\alpha_\pi}} \right)^2 \tau_s^{-1} + \left(\frac{\overline{\alpha_\pi}}{\overline{\alpha_s} + \overline{\alpha_\pi}} \right)^2 \tau_\pi^{-1} + \left(\frac{\overline{\alpha_n}}{\overline{\alpha_s} + \overline{\alpha_\pi}} \right)^2 \tau_n^{-1}. \tag{33}$$

$\tau_{\hat{p}}$ in Equation (33) corresponds to our measure of absolute price informativeness when there is a finite number of investors. There are two differences with respect to the baseline model presented in the main text. First, there are multiple sources of aggregate noise: the error of the private signal, ε_s; the error of he public signal, ε_π; and the demand of noise traders, n. Second, price informativeness is modulated by $\overline{\alpha_s} + \overline{\alpha_\pi}$ instead of by $\overline{\alpha_s}$ because there are two sources of external information about the fundamental θ.

A Bayesian external observer who only observes the price, learns from the price in the following way

$$\mathbb{E}\left[\theta | \hat{p}\right] = \frac{\tau_\theta \overline{\theta} + \tau_{\hat{p}} \hat{p}}{\tau_\theta + \tau_{\hat{p}}}.$$

Forecasting price efficiency (FPE) is then given by

$$\begin{aligned}
\mathcal{V}_{FPE} = \mathbb{V}ar\left(\mathbb{E}\left[\theta | \hat{p}\right]\right) &= \left(\frac{\tau_{\hat{p}}}{\tau_\theta + \tau_{\hat{p}}} \right)^2 \left(\tau_\theta^{-1} + \tau_{\hat{p}}^{-1} \right) \\
&= \left(\frac{\tau_{\hat{p}}}{\tau_\theta + \tau_{\hat{p}}} \right)^2 \left(\frac{\tau_\theta + \tau_{\hat{p}}}{\tau_\theta \tau_{\hat{p}}} \right) \\
&= \frac{\tau_{\hat{p}}}{\tau_\theta + \tau_{\hat{p}}} \tau_\theta^{-1}.
\end{aligned} \tag{34}$$

The expression for FPE in Equation (34) is the predicted variance of cash flows θ from prices. From this equation, it is easy to see that FPE confounds two effects. FPE can increase due to changes in the ability of prices to aggregate information,$\tau_{\hat{p}}$, or due to changes in the ease of forecastability of the fundamental, τ_θ. Hence, conditional on the variance of the fundamental remaining constant, FPE and price informativeness will co-move. However, without controlling for changes in fundamental volatility, one cannot make any inferences about price informativeness by looking at FPE.