ELSEVIER

Journal of Financial Markets 2 (1999) 273–305

Journal of FINANCIAL MARKETS

www.elsevier.nl/locate/econbase

Endogenous informed trading in the presence of trading costs: Theory and evidence☆

Jin-Wan Cho[a], Jhinyoung Shin[b], Rajdeep Singh[c,*]

[a]*The DuPree College of Management, Georgia Institute of Technology, USA*
[b]*School of Business Administration, Ajon University, South Korea*
[c]*Carlson School of Management, University of Minnesota, MN, USA*

Abstract

The basic premise of the model we propose is that market frictions (trading costs) force traders with market-wide information to strategically choose which securities to trade in. We study the effect of recognizing trading costs on the choices of informed traders and the resulting statistical properties of security prices. Specifically, we show that (1) stocks with intermediate β's have the *least* informative prices, even though they are traded by the greatest number of informed traders; (2) for high β securities, the contemporaneous correlation of prices is close to the correlation in fundamental values; (3) a security with a higher β, higher volume of liquidity trading and lower idiosyncratic variance is more likely to lead another security. With market capitalization as a proxy for the level of liquidity trading, these specific predictions of the model on the lead–lag relationship are also shown to be strongly supported by the data.

* Corresponding author.

E-mail addresses: jinwan.cho@mgt.gatech.edu (J.-W. Cho), fnshin@usthk.ust.hk (J. Shin), rajsingh@umich.edu (R. Singh)

☆We are especially indebted to Avanidhar Subrahmanyam (editor) and an anonymous referee for providing direction and invaluable insights. We also acknowledge the helpful comments of Sugato Bhattacharyya, Bruno Biais, Peter Bossaerts, David Brown, K.C. Chan, Rick Green, Larry Harris, Arnold Juster, Duane Seppi, Chester Spatt, Sheridan Titman and seminar participants at Carnegie Mellon University, Seoul National University, the Hong Kong University of Science and Technology, 1996 European Finance Association Conference at Oslo, 1997 Australian Finance and Banking Conference. We are grateful for the financial support from the William Larimer Mellon Fund and the Margaret and Richard M. Cyert Family Fund. The usual disclaimer applies.

1386-4181/99/$ - see front matter
PII: S1386-4181(99)00003-8

JEL classification: G20; G12

Keywords: Lead–lag effect; Market microstructure; Trading costs

1. Introduction

The evolution of security prices has attracted significant attention from a number of authors. Empirical studies have revealed some interesting patterns in the evolution of prices on both short-term and long-term basis. For instance, Lo and MacKinlay (1990b) establish that weekly returns on size-based portfolios exhibit a positive and asymmetric cross autocorrelation matrix. The empirical fact that large firms lead small firms cannot be fully explained by non-synchronous trading or thin markets. Chan (1992) studies the evolution of transaction prices. He finds an asymmetric lead–lag relation between futures and the cash index.[1]

This paper presents a theoretical framework in which we explicitly model trading costs and informed traders' decision to trade in a security. We show that these strategic choices made by informed traders impact the timing of information incorporation in prices. This not only helps us obtain predictions consistent with the empirically observed patterns mentioned above, it also helps in identifying characteristics of securities other than size, which might contribute to the lead–lag relationship.

We distinguish between two types of information in the context of financial markets. Information-based trading may occur due to unanticipated information either about the economy (i.e., a common factor shock) or about an event specific to a company (i.e., an idiosyncratic shock). We denote the dependence of a security on the common factor by β. Trading in some securities is more heavily influenced by common factor shocks while idiosyncratic shocks play a greater role for other securities. Thus, we model two types of informed traders: those with economy-wide information and others with company specific information. Trading costs allow us to endogenously determine the equilibrium number of informed traders of both types. The incorporation of the decision problem of informed traders regarding which securities to trade in also has important implications for the evolution patterns of prices. The statistical properties of security prices that emerge from a model incorporating such decision-making is markedly different from those exhibited by extant models. In particular, our model produces patterns quite similar to some that are observed in the data.

[1] There is strong evidence of the Major Market (MM) Index futures and the S&P 500 futures leading the MM cash index as well as its component stocks. On the other hand, the evidence for MM cash index leading the futures is weak.

We first study the informativeness of observed prices and establish a rather counter-intuitive result. On endogenizing the number of informed traders, we find that prices of securities with intermediate levels of dependence on the common factor (β) are least informative. This is true in spite of the fact that the total number of informed traders trading a particular security is maximized at this level of correlation. This result contrasts strongly with a constant level of informativeness across securities, which is obtained under the standard assumption of a fixed number of traders of each type. In our model, the number of informed traders are determined endogenously by a zero-profit no-entry condition. Thus, ceteris paribus an increase in the number of informed traders in equilibrium implies an increase in the amount of total profits made by the informed traders. This higher degree of adverse selection leads to less informative prices.

We next examine the contemporaneous correlation of observed security prices. Despite noise trading, for securities having a high β with the common factor, we find that the correlation of prices is quite close to the correlation in fundamental values. The higher correlation in observed prices is due to a larger number of informed traders with common factor information trading in these high β securities. In contrast, securities with low β have contemporaneous correlations much smaller than those of the fundamental values.

The most important implication of the model, we believe is about the lead–lag relationship amongst security returns. Previous work has shown that non-zero cross autocorrelations in short horizon returns can arise due to the partial adjustment of prices to information. We show that endogenization of traders' entry decisions in the face of trading costs implies that the autocorrelation matrix will be asymmetric. Specifically, we show that the cross autocorrelation amongst two securities' returns is monotonic and increasing in the leading security's β and the level of noise trading. The cross autocorrelation is also shown to be decreasing in the leading stock's idiosyncratic variance and the lagging stock's level of noise trading. Thus, difference in underlying fundamental parameters can give rise to an asymmetric cross autocorrelation matrix. Such an asymmetric lead–lag relation in security prices cannot be explained in the absence of trading costs when all informed traders with relevant information trade optimally.

We also provide empirical support for the model's predictions on the lead–lag relationship. If firm size is interpreted as a proxy for the level of liquidity trading then the above mentioned result is consistent with the observed lead–lag effect in size-based portfolio. We analyze the cross autocorrelations amongst 100 randomly chosen stocks and show that even for individual stocks, all else being equal, large firms lead small firms. More importantly, our empirical analysis provides support for the previously unknown prediction on the effect of β on the cross autocorrelations. We show that, as predicted by the theoretical model, the cross autocorrelation is increasing in the leading stock's β. The empirical analysis also provides strong support for the prediction that the cross autocorrelation is decreasing in the idiosyncratic variance of the leading firm. In short, all the unambiguous predictions of the model about the level of cross autocorrelation are supported by the empirical analysis.

Subrahmanyam (1991b) also analyzes a model similar to ours in which multiple trader types are allowed. His focus is on the introduction of basket trading in the presence of discretionary liquidity traders. Similar to our approaches, he, too, endogenizes the number of informed traders. However, in Subrahmanyam (1991b) the number of common factor traders is the same in all securities. We believe that traders with superior information about the common factor do not trade in all securities but, instead, strategically choose to trade in a subset of securities. Consequently, we choose a modeling strategy which identifies characteristics of individual securities that attract more (or less) common factor traders.[2] This modeling strategy leads to important differences in results from those in Subrahmanyam (1991b). We will make these differences clearer later in the paper.

Chan (1993) also shows that partial impounding of information in security prices results in non-zero cross autocorrelation in security price differentials. Assuming that information on high capitalization firms is more precise than that on low capitalization firms, he shows that large firms will lead small firms. In a model similar to the one developed in this paper, Cho et al. (1998) show that the cross autocorrelations amongst securities are also affected by the correlation in noise trading – the higher the correlation in noise trading the lower the level of cross autocorrelation. Bossaerts (1993), in a related paper, shows that informed trading based on market-wide information in all securities leads to a symmetric autocorrelation matrix. To get the empirically observed asymmetry, he is forced to restrict trading of the informed traders to specific portfolios. Instead of exogenously imposing an information structure or exogenously restricting the trading opportunities, we make assumptions that endogenize the strategic choices of informed traders in the presence of trading costs and get results consistent with the lead–lag relation.

The rest of the paper is organized as follows. The next section lays out the model. Section 3 solves for equilibrium and shows the strategic effects of one type of trader on the other. Section 4 studies the effects of endogenizing the number of informed traders on various statistical properties of prices and price differentials. Section 5 provides empirical support for the model. The last section concludes. All proofs are collected in the Appendix.

[2] Put another way, our model is targeted to capture the tension among the common factor traders and the security specific traders on a security by security basis, while Subrahmanyam (1991b) captures the effect on a market-wide basis.

2. Model

We consider a multi-period world with M securities, whose fundamental payoffs are driven by a single factor model. The payoffs of these are dependent not only on the common factor but also on the realizations of their idiosyncratic shocks. True payoffs of securities at the end of period t are given by

$$\tilde{\eta}_j^t = \tilde{\eta}_j^{t-1} + \Delta\tilde{\eta}_j^t = \tilde{\eta}_j^{t-1} + \beta_j\tilde{\theta}^t + \tilde{\varepsilon}_j^t \quad \text{for } j = 1, 2, \dots, M. \tag{1}$$

The random variables $\tilde{\theta}^t$'s, $\{\tilde{\varepsilon}_j^t\}_{j=1,\dots,M}$, are serially and mutually independent, and normally distributed with mean 0[3] and variances $(\sigma_\theta^2, \sigma_{\varepsilon_1}^2, \dots, \sigma_{\varepsilon_M}^2)$. We denote

$$\mathrm{Var}(\Delta\tilde{\eta}_j^t) = \beta_j^2\sigma_\theta^2 + \sigma_{\varepsilon_j}^2 \equiv \sigma_{\eta_j}^2. \tag{2}$$

There are four types of traders in our model: (1) informed traders with perfect information on the common factor; (2) informed traders in each security who only have perfect information about the idiosyncratic component of that particular security; (3) liquidity traders in each security who trade for reasons exogenous to the model; and (4) competitive market makers[4] in each security who set prices that give them zero expected profits conditional on their information. We assume that there are sufficiently large number of traders with private information on the common factor, and for each security the number of informed traders who have private access to the signals of idiosyncratic factor is also sufficiently large.[5] The basic structure of the market is similar to that in Kyle (1985) and Admati and Pfleiderer (1988). The demand by the liquidity traders, which gets aggregated with the demand by the informed traders, for each security is $\tilde{u}_i$ for $i = 1, \dots, M$, respectively. The random variables $\{\tilde{u}_i^t\}_{i=1,\dots,M}$ are all i.i.d. normal with mean 0 and variance $\sigma_{u_i}^2$ respectively. The liquidity demands are also independent of the random variables $\tilde{\theta}^t$ and $\{\tilde{\varepsilon}_i^t\}_{i=1,\dots,M}$. For each security, the competitive market maker observes the net trading order, determines the price as per the equilibrium price schedule and sells or buys the order imbalance.

[3] We make this assumption for simplification purposes. All our results go through if the securities had a positive mean.

[4] Although we do not explicitly model the game among market makers that guarantees zero profits for them, it can be thought of as a Bertrand competition of commitment to a price schedule before the start of the game.

[5] We later endogenously solve for the number of informed traders who participate in the trading of each security, and show that the equilibrium number of informed traders across securities depends on the individual characteristics of these securities. Thus, all comparative statics are based on the endogenously solved number of informed traders.

Market observes $\tilde{\theta}^{t-1}$ and $\tilde{\epsilon}_i^{t-1}$.

Informed traders observe realization of $\tilde{\theta}^t$ and $\tilde{\epsilon}_i^t$.

Market Makers announce price schedules.

Informed traders submit orders and liquidity trades are realized.

Prices at time t are set.

Market observes $\tilde{\theta}^t$ and $\tilde{\epsilon}_i^t$.

Fig. 1. Sequence of events.

The sequence of trading is as follows. At the beginning of period t, the informed traders who decide to participate in the trading pay the trading cost of T_c. Then, informed traders observe their respective signals and the market makers announce the price schedules for each security, which earns them zero expected profits for every realization, given their information. The informed traders take the price schedules as given and then place market orders to maximize expected profits. The orders of the informed traders are aggregated with the realized liquidity demand. The market makers observe the aggregate net orders, and then set prices according to the announced schedule and internalize the order imbalance. At the end of the period, the true payoffs are realized according to Eq. (1). The sequence of events is made clear in Fig. 1.

As mentioned earlier, in equilibrium[6] the market makers set the price at t equal to the expected value conditional on $\tilde{\eta}_i^{t-1}$ and the aggregate net trading order y_i^t. Using the linearity of the conditional expectation of normal variables, we define the price at t as

$$\begin{aligned} P_i^t &= E[\tilde{\eta}_i^t \mid \tilde{\eta}_i^{t-1}, y_i^t] \\ &= \tilde{\eta}_i^{t-1} + \lambda_i y_i^t \quad \text{for } i = 1, \dots, M. \end{aligned} \tag{3}$$

To complete the description of the market environment, we need to determine the number of informed traders in each security. Casual empiricism suggests that traders who trade on the basis of economy-wide information generally choose to trade only in securities which have a larger economy-wide component. Similarly, not many traders with economy-wide information trade in securities whose prices are mainly affected by idiosyncratic information. Instead of assuming the above comparative static on the number of informed traders, we obtain it

[6] We follow the literature in this area and solve for the unique linear equilibrium. This does not preclude the existence of other non-linear equilibria.

by assuming a trading cost structure. We assume that trading is a costly activity and the set-up cost of trading in each security is given by T_c, which needs to be spent by informed traders at every trading period. These trading costs[7] are different from the usual transaction costs as they are sunk costs, which do not depend upon the amount of trade. Although the assumed trading cost structure may seem somewhat arbitrary at first glance, it allows us to neatly close the model and the number of informed traders we get in each security is consistent with the above mentioned observation.[8]

Our binary choice cost structure might seem close to the often modeled cost of acquiring information;[9] one either spends the money to get the information or does not. The cost is independent of the quantity traded in equilibrium but the choice about whether to get information is not. Similarly, traders in our model decide to trade in specific securities and incur the relevant trading costs before the realization of the signal. These costs can be thought of as the costs of setting up the infrastructure to trade. The important difference from a modeling standpoint is that every trader in our model has to incur these costs for every security. However, traders in papers which model information acquisition cost will always choose to trade in all securities once they have acquired information on the common factor. We assume the trading cost structure instead of the explicit information acquisition structure because it helps us obtain a higher number of economy-wide informed traders in securities with a higher degree of the economy-wide component. In contrast an information acquisition structure will result in the same number of economy-wide informed traders in every security, which we believe is rather unrealistic.[10] For most of the paper we assume that the trading costs are the same across securities.

3. Number of informed traders in equilibrium

Security prices depend on the number of informed traders that choose to trade in the specific securities. In this section, we solve for the number of

[7] The results of this paper go through even if some additional marginal trading costs exist. What is important is the presence of the fixed set-up costs that force informed traders to choose which securities to trade in.

[8] An alternate assumption that the number of economy-wide informed traders in a security proportionally increases with the weight of the common factor in the security and the number of security-specific informed traders proportionally decreases will also obtain results similar to the ones in this paper.

[9] For example assumed in Subrahmanyam (1991b).

[10] The reader is reminded of the alternate assumption in footnote 8 which obtains almost all of the results.

informed traders in each security. We show that the informed traders of the same type, with common factor information or idiosyncratic information, are substitutes of each other. We also show that the traders of opposite types are complements. These allow us to make some empirically testable predictions about the number of analysts following stocks. We show that securities with intermediate levels of β will have the maximum total number of informed traders, including traders with economy-wide information.

Risk neutrality of informed traders implies that they will not trade in any other security where they do not have superior information. Thus, informed traders with idiosyncratic information will trade only in the security they have information about whereas informed traders with common factor information can potentially profit from trading in all securities.

Let $\{K_j\}_{j=1,\dots,M}$ be the number of informed traders with signal $\tilde{\theta}^t$, trading in security j and let $\{N_i\}_{i=1,\dots,M}$ be the number of informed traders with signal $\{\tilde{\varepsilon}_i^t\}_{i=1,\dots,M}$, trading in security i. The equilibrium price schedules in Eq. (4) are obtained by following the proof strategy in Subrahmanyam (1991b):[11]

$$\mathrm{P}_i^t = \tilde{\eta}_i^{t-1} + \lambda_i y_i^t \tag{4}$$

where

$$y_i^t = \frac{K_i}{\lambda_i(K_i+1)}\beta_i\tilde{\theta}^t + \frac{N_i}{\lambda_i(N_i+1)}\tilde{\varepsilon}_i^t + \tilde{u}_i,$$

$$\lambda_i = \frac{1}{\sigma_{u_i}}\sqrt{\frac{K_i}{(K_i+1)^2}\beta_i^2\sigma_\theta^2 + \frac{N_i}{(N_i+1)^2}\sigma_{\varepsilon_i}^2} \quad \text{for } i = 1,\dots, M.$$

Note that given the stationarity of all random variables the price schedules are identical for all t. The inverse of the price functional's slope, i.e., $1/\lambda_i$, is often used as a measure of market liquidity. A decrease in λ_i reduces the price impact of net trade, thereby increasing market liquidity. Informed traders take advantage of this increase in market liquidity by trading more aggressively and, consequently making higher trading profits.

Trading costs reduce the expected net profits of the informed traders. We assume that, in equilibrium, the number of informed traders adjust to a point where any increase in the number of informed traders will earn a negative expected profit net of trading costs.[12] Conditional on the magnitude of these trading costs and the security specific parameters, the number of informed traders for each security is endogenously determined to satisfy the zero profit condition of the informed traders. We assume that the number of traders is large enough to satisfy the zero profit condition in each security.

[11] Detailed proofs can be obtained from the authors on request.

[12] Essentially a no-entry condition.

The expected trading profits of the informed traders trading security i are a function of not only the number of informed traders but also the structure of fundamental payoff of the security. Thus, let $\pi_F^i(K_i, N_i, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i})$ be the expected trading profit of these informed traders with signal $\tilde{\theta}^t$ trading in security $i \in \{1, \dots, M\}$. Similarly, let $\pi_i(K_i, N_i, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i})$ be the expected trading profit of the informed traders with signal $\tilde{\varepsilon}_i^t$ trading in security $i \in \{1, \dots, M\}$. The profits of the informed traders conditional on the number of informed traders are given by

$$\pi_F^i(K_i, N_i, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i}) = \frac{\beta_i^2 \sigma_\theta^2}{(K_i + 1)^2} \frac{1}{\lambda_i}$$

$$= \frac{\beta_i^2 \sigma_\theta^2}{(K_i + 1)^2} \frac{\sigma_{u_i}}{\sqrt{\dfrac{K_i}{(K_i + 1)^2} \beta_i^2 \sigma_\theta^2 + \dfrac{N_i}{(N_i + 1)^2} \sigma_{\varepsilon_i}^2}}, \tag{5}$$

$$\pi_i(K_i, N_i, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i}) = \frac{\sigma_{\varepsilon_i}^2}{(N_i + 1)^2} \frac{1}{\lambda_i}$$

$$= \frac{\sigma_{\varepsilon_i}^2}{(N_i + 1)^2} \frac{\sigma_{u_i}}{\sqrt{\dfrac{K_i}{(K_i + 1)^2} \beta_i^2 \sigma_\theta^2 + \dfrac{N_i}{(N_i + 1)^2} \sigma_{\varepsilon_i}^2}}. \tag{6}$$

We assume that there exists at least one informed trader with information on $\tilde{\theta}$ and one with information on $\tilde{\varepsilon}_i$ for all i securities. Imposing this restrictions allows us to show that the profits of a set of informed traders are monotonic in the number of traders of the other kind.[13] Specifically we assume

$$\pi_F^i(1, 1, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i}) = \frac{\beta_i^2 \sigma_\theta^2}{2} \frac{\sigma_{u_i}}{\sqrt{\beta_i^2 \sigma_\theta^2 + \sigma_{\varepsilon_i}^2}} > T_c, \tag{7}$$

$$\pi^i(1, 1, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i}) = \frac{\sigma_{\varepsilon_i}^2}{2} \frac{\sigma_{u_i}}{\sqrt{\beta_i^2 \sigma_\theta^2 + \sigma_{\varepsilon_i}^2}} > T_c. \tag{8}$$

[13] We thank the referee for pointing out that for an increase in the number of informed traders from 0 to 1, in fact, reduces the profits of the informed traders of the other kind. We should also point out that the result also depends on the assumption, standard in the literature, about the risk neutrality of informed traders. Subrahmanyam (1991a) has shown that if informed traders are risk averse then λ is not necessarily monotonic in the number of informed traders. Thus, in the case of extreme risk aversion the complementarity result may break-down.

Trading profits, in security i, of the informed traders with common factor information ($\tilde{\theta}^t$) are increasing in β_i. This is, obviously, due to the increase in the relative advantage of their information in the sense that a greater portion of the true payoff is determined by the common factor which they have relevant information about. The reverse is true for the traders with idiosyncratic factor $\tilde{\varepsilon}_i^t$. An increase in the level of liquidity trading, σ_{u_i}, has positive effect on the trading profits earned by both types of informed traders due to the increases in market liquidity. Not surprisingly, the expected trading profits of the informed traders, due to increased competition, is decreasing in the number of informed traders with the same signal. Similar to the usage in Bulow et al. (1985), traders with different information in our model are 'complements' in that the profits of the informed are increasing in the number of informed traders with a different signal.[14] The above comparative statics are collected in the next lemma.

Lemma 1. *Expected trading profits earned from trading security i has the following properties*:

1. $\pi_F^i(K_i, N_i, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i})$ *increases in* N_i, β_i *and* σ_{u_i} *while decreases in* K_i *and* σ_{ε_i}.
2. $\pi_i(K_i, N_i, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i})$ *increases in* K_i, σ_{ε_i} *and* σ_{u_i} *but decreases in* N_i *and* β_i.

As the profits of informed traders are monotonic in the number of informed traders, we can solve for the equilibrium number of informed traders in each market that satisfy their zero profit condition net of trading costs. We compute, for a given level of trading costs, the equilibrium number of traders trading in a particular security. Let the number of traders with the common factor trading in security i be K_i^* for $i = 1, \dots, M$. Also let N_i^* be the number of traders with the idiosyncratic component information trading in security i for $i = 1, \dots, M$.

The number of informed traders in the securities are affected not only by the level of trading costs but also by the number of informed traders of the opposite kind. The equilibrium number of traders with information on the idiosyncratic component will favorably affect the number of informed traders with common factor information and vice versa. Thus, we need to simultaneously solve the problems of both kinds of traders to get the number of traders in equilibrium. With an assumed constant level of trading costs T_c for each security, the number of traders is determined by solving the system of equations in Lemma 2.

[14] This result is not due to the assumed trading cost structure. Subrahmanyam (1991b) gets a similar comparative static in a model with information acquisition costs.

Lemma 2. *Given T_c, σ_{u_i}, σ_{ε_i}, β_i such that inequalities* (7) *and* (8) *are satisfied, the equilibrium K_i^* and N_i^* simultaneously solve*:[15]

$$\pi_F^i(K_i^*, N_i^*, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i}) = T_c, \tag{9}$$

$$\pi_i(K_i^*, N_i^*, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i}) = T_c. \tag{10}$$

Comparative statics on the number of informed traders in equilibrium, which follow from the properties of informed trader's profits presented in Lemma 1, are formalized in the following proposition.

Proposition 1. *Equilibrium number of traders with information on*:

1. *the common factor increases in β_i and σ_{u_i} but decreases in σ_{ε_i} and T_c; and*
2. *the idiosyncratic factor increases in σ_{ε_i} and σ_{u_i} but decreases in β_i and T_c.*

The increase in the trading costs of one group of informed traders adversely affects the other group's expected trading profit and actually reduces the number of informed traders in both groups. As one group of informed traders collectively takes a less aggressive trading strategy due to the higher trading cost or less accurate information, the price moves more sensitively to the net trading order. This change in the price schedule adversely affects the expected trading profit of the other group of informed traders, and they collectively take a less aggressive trading strategy.

The results in this section allow us to predict that: (1) The total number of informed traders (analysts) trading in a stock, for a given level of variance, should be the highest for stocks with intermediate levels of β;[16] (2) Stocks with very low β may not have any market-wide analysts trading in them and, thus, will not attract the attention of program traders, et cetera; and (3) If the number of market-wide analysts trading in a stock increases due to some exogenous reasons like a decrease in the trading cost, then the number of analysts following that particular stock should also increase, even if their trading costs remain constant. We are not aware of any study that examines the above predictions. As it is hard to distinguish between information based trading and liquidity based trading, we believe it will be quite hard to find any direct evidence supporting these predictions. However, the next section develops implications for observed

[15] We essentially ignore the integer problem and assume that the number of traders is continuous. Even when the integer problem is explicitly considered, we do not expect any major change in the results. For instance, if the integer problem is taken into account in determining the equilibrium number of informed traders, then K_i^* and N_i^* satisfy $\pi_F^i(K_i^*, N_i^*, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i}) \geqslant T_c > \pi_F^i(K_i^* + 1, N_i^*, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i})$ and $\pi_i(K_i^*, N_i^*, \beta_i, \sigma_{\varepsilon}, \sigma_{u_i}) \geqslant T_c > \pi_i(K_i^*, N_i^* + 1, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i})$.

[16] Please refer to the Proof of Proposition 2 for the formal proof of this statement.

prices, given the number of informed traders. Testing of these implications, we believe, will also provide an indirect support for the results in the current section.

4. Implications of endogenizing the number of informed traders

In this section, we use the equilibrium number of informed traders to derive the time series and cross sectional properties of securities' prices. Later in the paper we will empirically test these predictions by analyzing a cross-section of securities.

4.1. Informativeness of prices

We begin by studying the amount of information impounded into the prices due to the trading of the informed. At the beginning of period t, $\tilde{\eta}_i^{t-1}$ is publicly known and we can see that the price $(\lambda_i y_i^t)$ reveals some of the information on $\Delta\tilde{\eta}_i^t$ to the market. The degree to which the informed traders' private information is revealed in the price, as a function of the various parameters, can be measured by the conditional variance of $\Delta\tilde{\eta}_i^t$. However, as the variance of the fundamental payoff of each security is potentially different, comparing the conditional variance across securities with different fundamental variance is not very meaningful. Thus, for the purpose of Section 4.1 we choose to compare the informativeness of prices of securities with the same fundamental variance, i.e., for a fixed σ_η. More specifically, we have a set of securities such that $\beta_h^2\sigma_\theta^2 + \sigma_{\varepsilon_h}^2 = \sigma_\eta^2$ for all h, and among the securities in this set, if $\beta_j > \beta_k$ then $\sigma_{\varepsilon_j}^2 < \sigma_{\varepsilon_k}^2$ holds.

Proposition 1 shows that the number of traders with common factor information and those with idiosyncratic component information who find it optimal to trade in the particular security will depend on the structure of fundamental payoffs. Consequently, in equilibrium for a fixed level of trading costs, prices will reflect different levels of information about the common factor and idiosyncratic factor depending on the relative weight of the common factor and the idiosyncratic shock. In the following proposition, we show that among the securities with identical σ_η, the informativeness of price is the least for a security with an intermediate level of β.

Proposition 2.

1. *If trading costs are greater than zero, then*

$$\mathrm{Var}(\Delta\tilde{\eta}_i^t \mid \lambda_i y_i^t) = \sigma_\eta^2\left[1 - \left(\frac{K_i^*}{K_i^*+1}\,\frac{\beta_i^2\sigma_\theta^2}{\beta_i^2\sigma_\theta^2+\sigma_{\varepsilon_i}^2} + \frac{N_i^*}{N_i^*+1}\,\frac{\sigma_{\varepsilon_i}^2}{\beta_i^2\sigma_\theta^2+\sigma_{\varepsilon_i}^2}\right)\right]. \tag{11}$$

2. *For a given* σ_η, $\mathrm{Var}(\Delta\tilde{\eta}_i^t \mid \lambda_i y_i^t)$ *is concave and non-monotonic in* β_i *and achieves a maximum at intermediate level of* $\beta_i = \hat{\beta}$ *where* $\hat{\beta} = \sigma_\eta/\sqrt{2}\sigma_\theta$.
3. *The total number of traders* $N^*(\beta_i) + K^*(\beta_i)$ *is also maximized at* $\beta_i = \hat{\beta}$.

One might expect that the informativeness of prices would be higher as the total number of traders increases. In fact, the opposite is true. Given σ_η, the total number of informed traders is maximum for $\beta_i = \sigma_\eta/\sqrt{2}\sigma_\theta$, but surprisingly the conditional variance of prices is also the highest at the same point. To obtain intuition for the above result it is instructive to examine the standard Kyle (1985) model in which the security fundamentals consist of only one random variable and there are $L \geqslant 2$ informed traders. Now keeping the variance of the fundamentals constant, a decrease in the number of informed traders leads to an increase in the total profits for the informed traders and a decrease in the informativeness of prices. All else being equal, the higher the profits made by the informed traders due to a change in their numbers, the higher the adverse selection component and the lower the informativeness of prices.[17]

The total profit of informed traders in our model is determined by the free entry condition and is equal to the total number of informed traders times the magnitude of the trading cost.[18] The number of traders in each security is given by the zero profit condition for each informed trader after accounting for the trading costs. Thus, an increase in the number of informed traders in our model implies that there is more profit being made at the expense of liquidity traders. At the intermediate level of β, the number of informed traders and the profits made by informed traders are maximized. This is because the complementarity effect is the strongest when the two sources of information are equally weighted. Consequently, the highest level of adverse selection leads to the lowest level of price informativeness at $\hat{\beta}$.

4.2. Co-Movement of security prices

We next investigate a measure of co-movement in security prices. We show that the securities with a high dependence on the common factor, β, will have price innovations which are almost as highly correlated as the fundamentals. In contrast, securities with a small β will have contemporaneous correlations of prices much smaller than that of fundamental values. For the analyses of the following two subsections, we use the price differential of security i at period

[17] The increase in profits does not necessarily imply a decrease in informativeness. This is especially true for the limit case, when the number of informed traders goes from 0 to 1 both the profits and the informativeness increases. We thank the referee for pointing out this caveat.

[18] In contrast, an *exogenous* increase in the number of informed traders in most models implies a decrease in the total profits made by the informed traders and, thus, leads to an increase in the informativeness of prices.

t from Eq. (4).

$$\begin{aligned}\Delta P_i^t &= P_i^t - P_i^{t-1}\\ &= \beta_i\tilde{\theta}^{t-1} + \tilde{\varepsilon}_i^{t-1} + \frac{K_i}{K_i+1}\beta_i\tilde{\theta}^t + \frac{N_i}{N_i+1}\tilde{\varepsilon}_i^t - \frac{K_i}{K_i+1}\beta_i\tilde{\theta}^{t-1}\\ &\quad - \frac{N_i}{N_i+1}\tilde{\varepsilon}_i^{t-1} + \lambda_i\tilde{u}_i^t - \lambda_i\tilde{u}_i^{t-1}\\ &= \frac{K_i}{K_i+1}\beta_i\tilde{\theta}^t + \frac{N_i}{N_i+1}\tilde{\varepsilon}_i^t + \frac{1}{K_i+1}\beta_i\tilde{\theta}^{t-1}\\ &\quad + \frac{1}{N_i+1}\tilde{\varepsilon}_i^{t-1} + \lambda_i\tilde{u}_i^t - \lambda_i\tilde{u}_i^{t-1}. \end{aligned} \tag{12}$$

From Eq. (12), we have

$$\begin{aligned}\mathrm{Var}(\Delta P_i^t) &= \frac{K_i^2+1}{(K_i+1)^2}\beta_i^2\sigma_\theta^2 + \frac{N_i^2+1}{(N_i+1)^2}\sigma_{\varepsilon_i}^2 + 2\lambda_i^2\sigma_{u_i}^2\\ &= \beta_i^2\sigma_\theta^2 + \sigma_{\varepsilon_i}^2 = \sigma_{\eta_i}^2. \end{aligned} \tag{13}$$

The second equality is derived from equilibrium λ_i in Eq. (4). To compare the co-movements among securities we normalize the correlation in price changes with the correlation amongst fundamentals and call it the relative correlation of price changes, defined as

$$\frac{Corr(\Delta P_i^t, \Delta P_j^t)}{Corr(\Delta\tilde{\eta}_i^t, \Delta\tilde{\eta}_j^t)} \tag{14}$$

Because of trading costs, the number of traders in equilibrium will depend on the weight of the common factor in a security's fundamentals and, consequently, the correlation in prices will also be its function. Now if β_i is high or if σ_{ε_i} is low, then traders with the idiosyncratic component information trade less aggressively and the realization of the idiosyncratic factor will have a much smaller effect on prices than on the fundamental values. Consequently, the correlation in prices will be closer to the correlation in fundamental values for high β_i or low σ_ε securities. Of course, the correlation of prices cannot be higher than that of fundamental values because of the noise in prices due to independent liquidity trading. The next proposition summarizes the result.

Proposition 3. *For positive trading costs,*

$$\frac{Corr(\Delta P_i^t, \Delta P_j^t)}{Corr(\Delta\tilde{\eta}_i^t, \Delta\tilde{\eta}_j^t)} = \frac{K_i^* K_j^* + 1}{(K_i^*+1)(K_j^*+1)} \tag{15}$$

which increases in β_i, β_j, σ_{u_i} *and* σ_{u_j} *but decreases in* σ_{ε_i}, σ_{ε_j} *and* T_c.

As is shown in Proposition 3, contemporaneous correlation between price differentials are critically dependent on the trading activities of traders with common factor information. As they trade a particular security more aggressively because its fundamental payoff is highly correlated with common factor or it has high variance of liquidity trading, then its price's contemporaneous correlation with other securities gets close to the correlation between fundamental payoffs. Since higher trading costs deter any trading activities by informed traders, naturally it has adverse impact on contemporaneous correlation.

4.3. Effect on cross autocorrelations of short horizon returns

Subrahmanyam (1991b) and Chan (1993) have shown that, if payoffs are generated by a common factor model, then the short horizon price differentials will be cross autocorrelated. Cho et al. (1998) show that this result is more general than previously thought. They show that the common factor is neither necessary nor sufficient. In this subsection, we show that price differentials will also be cross autocorrelated in current model. More importantly, the next proposition shows that the level of cross autocorrelation is a function of the various parameters of the model and cross autocorrelation matrix is, in general, asymmetric. This allows us to empirically test the validity of the model.

Proposition 4. *For $i, j = 1, \ldots, M$ and $i \neq j$*

1. *the cross autocorrelations of price differentials are given by,*

$$Corr(\Delta P_i^t, \Delta P_j^{t+1}) = \frac{K_i^*}{K_i^*+1}\frac{1}{K_j^*+1}Corr(\Delta\tilde{\eta}_i^t, \tilde{\theta}^t)Corr(\Delta\tilde{\eta}_j^{t+1}, \tilde{\theta}^{t+1})$$

$$= \frac{K_i^*}{K_i^*+1}\frac{1}{K_j^*+1}\frac{\beta_i\sigma_\theta}{\sqrt{\beta_i^2\sigma_\theta^2+\sigma_{\varepsilon_i}^2}}\frac{\beta_j\sigma_\theta}{\sqrt{\beta_j^2\sigma_\theta^2+\sigma_{\varepsilon_j}^2}}. \qquad (16)$$

2. *$Corr(\Delta P_i^t, \Delta P_j^{t+1})$ increases in β_i and σ_{u_i} but decreases in σ_{ε_i} and σ_{u_j}.*

As can be seen from Eq. (16), the cross autocorrelations in the above proposition are in general non-zero. This is because the innovations of the common factor are only partially revealed in the prices at t and are fully impounded into prices only at the later period as $\Delta\tilde{\eta}_i^t$ is fully revealed at the end of period t. For example, if traders with common factor information receive a positive signal, then they will submit a buy order in both securities. On average, a fraction of his information will be impounded in the price change at time t and the rest of the innovation will be reflected in the price change of the following period. Thus, ceteris paribus, if the price of security i went up in period t, then the price of security j is also likely to go up in the following period.[19] Thus price innovation of one security, on average, has information on the correction in the other security's price in the next period. This information gives rise to positive cross autocorrelations.

The level of cross autocorrelation as derived in Eq. (16) is a function of the specific parameter values of each security. This is especially important as it allows us to test the model. We know from the work of Lo and MacKinlay (1990b) and Chan (1993) that returns of portfolios not only show significant cross autocorrelation but also these cross autocorrelations are asymmetric – returns of large firms lead the returns of the small firms. This relation has been referred to as the lead–lag anomaly.

Endogenizing, the number of informed traders in each security allows for a natural asymmetry in the autocorrelation matrix. That is, in general, $Corr(\Delta P_i^t, \Delta P_j^{t+1}) \neq Corr(\Delta P_j^t, \Delta P_i^{t+1})$. As can be seen from the statement of the proposition, the only way for the cross autocorrelation amongst security i and j to be symmetric is if $K_i^* = K_j^*$. This would be the case if all traders with common-factor information chose to trade in every security. This lack of strategic choice will imply a symmetric cross autocorrelation matrix. However, in our model K^* depends on the fundamental parameters of the security. Consequently, securities which have a higher K^* will lead securities which have a lower K^* and the resulting cross autocorrelation matrix will be asymmetric.

The second part of the proposition has provided the comparative statics on the cross autocorrelation, which we will take to the data in the next section. The comparative static exercise is useful as it allows us to predict the characteristics of securities which are more likely to lead.[20] For example, Proposition 4 predicts that, all else being equal, securities with a higher level of liquidity trading will lead securities with a lower level of liquidity trading. A higher level of liquidity trading increases the number of informed traders with common factor information (K^*) in equilibrium, which increases the level of cross autocorrelation. It has been suggested that large firms attract a higher level of uninformed non-strategic trading. If that is true then our result would be consistent with the empirically observed fact that large firms lead small firms.[21] Thus, the model potentially provides a theoretical underpinning for the observed lead–lag relationship.

[19] However, the price changes of individual securities will be serially uncorrelated. This is because the pricing errors in period 1 due to liquidity trading and noisy information of the informed will be exactly offset in the following period. That is, the error and the correction are perfectly negatively correlated, which is not the case across two different securities.

[20] Chan (1993) also identifies a condition under which one security will lead another. He shows that securities which have traders with more precise information about their payoff lead other securities. Chan conjectures that larger firms will have traders with more precise information.

[21] We thank the referee for this insight.

Proposition 4 not only provides results consistent with previously observed phenomenon but also provides new testable implications. We have shown that the stocks that lead are more likely to have a higher β and a lower idiosyncratic variance. The cross autocorrelation $Corr(\Delta P_i^t, \Delta P_j^{t+1})$ is increasing in β_i as higher β_i leads to an increase in the number of factor informed traders, which implies that more of the common factor innovation of time t is incorporated in the price of security i at time 1. Thus, $Corr(\Delta P_i^t, \Delta P_j^{t+1})$ is increasing in β_i. However, an increase in β_j has two opposing effects. A higher β_j implies that in general the security has more common factor dependence and thus the price adjustment at time $t+1$ will be correlated to the change in price of security i at time t, which will tend to increase $Corr(\Delta P_i^t, \Delta P_j^{t+1})$. On the other hand, an increase in β_j results in an increase in factor informed traders in security j. Thus, little of the common factor of time t remains to be incorporated in the price of security j at time $t+1$, which will tend to decrease $Corr(\Delta P_i^t, \Delta P_j^{t+1})$. Similar intuition holds for the case of the idiosyncratic variance.

In this subsection, we have shown that cross autocorrelations will be symmetric if we do not explicitly recognize the existence of trading costs. Recognizing trading costs leads to a prediction of asymmetry in the autocorrelation matrix. The predicted asymmetry is not only broadly consistent with the observed lead–lag relationship but also provides further testable implications. The empirical support of these predictions is provided in the next section.

5. Empirical evidence

In this section, we provide evidence supporting some of the theoretical predictions of the model. The sharpest, and in our opinion the most interesting, prediction of the model is on the asymmetry of cross autocorrelations of stock returns. Previous empirical studies have shown that returns on a portfolio of large firms lead the returns of a portfolio of smaller firms (see for example, Lo and MacKinlay, 1990b). Previous attempts at providing explanations for these *asymmetric* cross autocorrelations have been based on market illiquidity of small stocks (Lo and MacKinlay, 1990a) and assumed better signals in large stocks (Chan, 1993). We are not aware of any study which has empirically investigated the relationship between the lead–lag behavior and the security's fundamental variables.

In contrast, in the theoretical model presented in this paper we have shown that the asymmetry of cross autocorrelations can arise from differences in underlying fundamental parameters. Specifically, Proposition 4 has shown that the level of cross autocorrelation amongst two stocks is a function of six exogenous parameters, some of which might be correlated with the size of the firm and the others clearly are not. The proposition, thus, provides us with a natural way to test the validity of our predictions against explanations which just depend on firm size.

Table 1
Summary statistics

This table reports the summary statistics of the variables that are used in the empirical analysis. First, we select 100 NYSE firms randomly, for which data are available for the entire period from 1990 to 1994 in the Center for Research in Security Prices (CRSP) data base. Then, the daily returns together with the daily returns on S&P 500 index are obtained for the sample period. *CAC* is the cross autocorrelations using daily returns, and there are 9900 (100×99) pair-wise CACs for a given sample period. β is the β coefficient between the individual stock and the S&P 500 index. For each of the 100 stocks, β is estimated via the market model using weekly returns for the given sample period. Third, σ_{ε_i} is the standard deviation of the idiosyncratic shocks to the fundamental payoffs of an individual stock, and for each of the 100 stocks, σ_{ε_i} is estimated by using Eq. (13) for the sample period. Last, Mkcap is the market capitalization in \$1 billion at the beginning of each sample year. For the whole period sample, *Mkcap* is estimated by taking an average of 5 yearly market capitalizations. Since Mkcap is a firm-specific variable, there are 100 Mkcaps for a sample period.

Variables	Statistic	1990	1991	1992	1993	1994	Whole period
	mean	4.29%	3.64%	1.21%	1.38%	1.69%	2.64%
CAC	stdev	7.76%	6.77%	6.54%	6.59%	6.72%	3.49%
	max	36.49%	32.27%	29.97%	25.71%	23.77%	19.78%
	min	− 25.68%	− 22.19%	− 26.26%	− 24.14%	− 24.00%	− 9.69%
	mean	0.831	0.872	0.768	0.578	0.379	0.814
β	stdev	0.490	0.563	0.594	0.522	0.377	0.403
	max	2.174	2.503	2.317	2.431	2.012	1.934
	min	− 0.216	− 0.442	− 0.794	− 0.707	− 0.605	0.014
	mean	0.019	0.020	0.018	0.017	0.016	0.018
σ_{ε_i}	stdev	0.010	0.013	0.010	0.010	0.009	0.009
	max	0.065	0.082	0.057	0.071	0.073	0.059
	min	0.006	0.004	0.004	0.004	0.005	0.006
	mean	1.368	1.228	1.559	1.701	1.931	1.564
Mkcap	stdev	2.252	2.048	2.571	2.630	2.800	2.375
	max	10.968	10.063	13.691	14.392	11.935	10.945
	min	0.007	0.003	0.008	0.010	0.010	0.008

Proposition 4 has shown that the cross autocorrelation between price change of stock i (leading stock) at time t with the price change of stock j (following stock) at time $(t+1)$ is a function of the following parameters: (i) the dependence of the stocks' fundamentals on the common factor θ as measured by β_i and β_j; (ii) the weight of the idiosyncratic component in both stocks' as measured by $\sigma^2_{\varepsilon_i}$ and $\sigma^2_{\varepsilon_j}$; and (iii) intensity of liquidity trading in both securities represented by $\sigma^2_{u_i}$ and $\sigma^2_{u_j}$. More importantly the proposition predicts that the cross autocorrelation is monotonic and increasing in β_i and $\sigma^2_{u_i}$ and is decreasing in $\sigma^2_{\varepsilon_i}$ and $\sigma^2_{u_j}$, where i is the leading security. The theoretical model, however, is unable to predict a monotonic relationship between cross autocorrelation and the other two

Table 2
Firm characteristics and cross autocorrelation

This table reports the regression results from the model:

$$CAC_{ij} = \alpha + \gamma_1\beta_i + \gamma_2\beta_j + \gamma_3\sigma_{\varepsilon_i} + \gamma_4\sigma_{\varepsilon_j} + \gamma_5\,Mkcap_i + \gamma_6\,Mkcap_j.$$

The last column reports the results for the entire sample period while the other columns present results from the separate estimations on the five sub-periods. In the model CAC_{ij} is the cross autocorrelation from individual stock returns when stock i is assumed to lead stock j; β_i is the β coefficient of the leading stock with respect to the S&P 500 index, while β_j is the β coefficient of the lagging stock; σ_{ε_i} is the expected size of the idiosyncratic shocks to the leading stock, while σ_{ε_j} is that to the lagging stock; $Mkcap_i$ is the market capitalization of the leading stock i at the beginning of the sample period, while $Mkcap_j$ is that of the lagging stock j. For the whole period sample case, however, *Mkcap* is the average of the 5 yearly market capitalizations. In the actual estimation *Mkcap* is in \$10 billion. Market capitalization is used to proxy the level of liquidity trading in the individual stocks. We conjecture that the liquidity trading in more intensive for a stock with a larger capitalization. The theoretical model predicts γ_1 and γ_5 to be positive, and γ_3 and γ_6 to be negative. For each sample period, there are 9900 (100×99) observations. In the parentheses, the t-statistics using the Heteroskedasticity corrected standard errors are reported.

	1990	1991	1992	1993	1994	Whole period
Intercept	0.016[c]	0.009[c]	0.004[b]	0.017[c]	0.017[c]	0.012[c]
	(6.65)	(4.67)	(1.95)	(8.36)	(8.06)	(10.16)
β_i	0.052[c]	0.022[c]	0.008[c]	0.014[c]	0.022[c]	0.028[c]
	(32.04)	(18.02)	(6.94)	(10.60)	(12.33)	(35.60)
β_j	0.034[c]	0.25[c]	0.010[c]	0.004[c]	0.002	0.020[c]
	(21.62)	(20.44)	(8.43)	(2.84)	(0.87)	(22.46)
σ_{ε_i}	− 1.026[c]	− 0.516[c]	− 0.405[c]	− 0.364[c]	− 0.382[c]	− 0.705[c]
	(− 13.61)	(− 9.73)	(−5.89)	(− 5.11)	(− 5.27)	(− 19.02)
σ_{ε_i}	− 1.198[c]	− 0.312[c]	0.124[a]	− 0.403[c]	− 0.235[c]	− 0.571[c]
	(− 16.32)	(− 5.68)	(1.79)	(− 6.20)	(− 3.20)	(− 16.80)
$Mkcap_i$	0.027[c]	0.016[c]	0.002	0.002	0.008[c]	0.007[c]
	(7.81)	(5.10)	(0.61)	(0.18)	(2.95)	(4.30)
$Mkcap_j$	− 0.042[c]	− 0.000	− 0.007[c]	− 0.003	− 0.005[a]	− 0.016[c]
	(− 13.28)	(− 0.16)	(− 2.64)	(− 0.31)	(− 1.80)	(− 10.86)

[a]Significant at 10%.
[b]Significant at 5%.
[c]Significant at 1%.

parameters, β_j and $\sigma^2_{\varepsilon_j}$. In the following, we perform a cross-sectional regression analysis on pair-wise cross autocorrelations between individual stocks as a function of the above parameters. We hope to achieve two goals. First, we would like to test the predictions regarding the directional relationship between cross autocorrelation and parameters for which the theoretical model predicts a definite relationship. Second, we hope to empirically learn about the direction of the relationship between cross autocorrelation and the other parameters for which the theory model is silent.

We study the impact of the underlying parameters on daily cross autocorrelations of individual stocks. The data used for analysis is obtained as follows: We randomly select 100 NYSE firms from the Center for Research in Security Prices (CRSP) database. We do that by running a random number generator on the permanent numbers of the stocks. The list of selected sample firms is provided in Table 3. Once the sample firms are selected, we obtain their daily returns and also the daily return on the S&P 500 index for the period from 1990 to 1994. We also obtain their market capitalization at the beginning of every year in our sample period.

Before we can report on our empirical analysis we need to discuss the choice of our dependent and independent variables. In this cross sectional study our dependent variable is simply the pair-wise cross autocorrelation daily returns of the 100 chosen securities.[22] Specifically, it is the correlation between the daily returns of security *i* with one period lagged daily returns of security *j*. The cross correlations are calculated for the whole sample period and separately for the individual years to be treated as sub-periods. Since there are 100 stocks, for each sample period, we obtain 9900 ($=100\times 99$) pair-wise cross autocorrelations. These observations will constitute our dependent variable's series for the cross sectional regression analysis.

To obtain the independent variable β for every stock we run a market model using weekly returns of the stock and the S&P 500 index. We have, thus, chosen the S&P 500 as our underlying common factor for all securities. As β_i denotes the relationship between the change in fundamentals of stock *i* with that of the index, it is inappropriate to obtain this variable from daily return series. This is because, we believe, that the daily return series is more prone to the effect of noise trading and other market microstructure frictions. The relationship obtained from the weekly return series should be a better proxy for the underlying fundamental relationship.[23] The market model is estimated using data from the

[22] In contrast, most previous work aggregates the cross autocorrelations across portfolios. For example, in Chan (1993) the cross autocorrelations between individual securities and index are averaged over size-stratified portfolios. In this paper, however, we do not work with the averages of pair-wise cross autocorrelations for the following two reasons. First, when the cross autocorrelations are averaged, the effects of idiosyncratic factors, such as σ_ε, on cross autocorrelation may well be diversified away, rendering the empirical analysis of these effects problematic. Second, given the predictions of the theory model, we need to conduct a multivariate analysis. In a multivariate analysis, unlike in a univariate case, working with averages can be quite cumbersome. Hence we choose to analyze the data at the individual security level.

[23] Of course, our results are joint tests of the model and the specific choice of the proxy variables.

Table 3
List of selected sample firms

The following table lists the firms randomly selected for the empirical analysis. The market capitalization is the average of the five market capitalizations at the beginning of 1990–1994.

Perm. number	Ticker symbol	Company name	Market capitalization
10065	ADX	Adams Express Co.	0.576
10145	ALD	Allied Signal Inc.	6.947
10302	CY	Cypress Semiconductor Corp.	0.444
11340	CG	Columbia Gas Sys Inc.	1.538
11762	ETN	Eaton Corp.	2.478
12570	ITT	I T T Corp.	7.874
12650	KSU	Kansas City Southn Inds Inc.	0.936
13936	PVH	Phillips Van Heusen Corp.	0.461
15368	WX	Westinghouse Electric Corp.	6.993
16396	BKF	Baker Fentress & Co.	0.325
17830	UTX	United Technologies Corp.	6.635
18032	SPP	Scott Paper Co.	2.933
18286	RAY	Raytech Corp. De.	0.008
18542	CAT	Caterpillar Inc.	5.929
18956	TY	Tri Continental Corp.	1.568
18964	PEO	Petroleum & Resources Corp.	0.283
19502	WAG	Walgreen Co.	4.237
19895	MEA	Mead Corp.	2.180
19916	VO	Seagram Ltd.	9.468
20248	CPY	C P I Corp.	0.368
21231	WGL	Washington Gas Lt Co.	0.703
22509	PPG	P P G Industries Inc.	5.958
22840	SLE	Sara Lee Corp.	10.945
24248	MPL	Minnesota Power & Light Co.	0.908
24870	CIP	C I P S C O Inc.	0.912
25080	UPK	United Park City Mines Co.	0.011
26331	ADP	Allied Products Corp De.	0.045
26454	SNS	Sundstrand Corp.	1.304
27780	ST	S P S Technologies Inc.	0.129
28847	U	Usair Group Inc.	0.837
33785	LOR	Loral Corp.	1.551
36468	SHW	Sherwin Williams Co.	2.266
36469	FTU	First Union Corp.	4.063
36469	FTU	First Union Corp.	4.063
38420	CNK	Crompton & Knowles Corp.	0.805
41081	SNC	Shawmut National Corp.	1.278
41136	KZ	Kysor Industrial Corp De.	0.068
41179	NLC	Nalco Chemical Co.	2.332
42893	HTD	Huntingdon International Hdg Plc.	0.306
43481	SXI	Standex International Corp.	0.277
44433	UCO	Union Corp.	0.126
45487	NSH	Nashua Corp.	0.208

Table 3. Continued.

Perm. number	Ticker symbol	Company name	Market capitalization
45495	RDC	Rowan Companies Inc.	0.686
46181	J	Jackpot Enterprises Inc.	0.073
46228	RB	Reading & Bates Corp.	0.210
47837	MKC	Marion Merrell Dow Inc.	7.938
48347	LZB	La Z Boy Chair Co.	0.442
48506	DNB	Dun & Bradstreet Corp.	9.509
49701	AUG	Augat Inc.	0.228
52038	SYY	Sysco Corp.	4.155
52396	SUP	Superior Industries Intl Inc.	0.496
53364	SP	Spelling Entertainment Group Inc.	0.349
53372	IRT	I R T Property Co.	0.178
54279	DSL	Downey Savings & Loan Assn.	0.251
54471	SFS	Super Food Services Inc.	0.160
55597	BW	Brush Wellman Inc.	0.253
56354	SZ	Sizzler International Inc.	0.348
58166	NSB	Northeast Federal Corp.	0.027
59301	HRD	Hannaford Brothers Co.	0.813
59360	LGN	Logicon Inc.	0.127
59897	KUH	Kuhlman Corp.	0.076
59926	MNR	Manor Care Inc.	1.010
60521	KII	Keystone International Inc.	0.865
60986	NWL	Newell Company	2.416
61284	BBI	Barnett Banks Inc.	2.797
61313	DCI	Donaldson Inc.	0.396
61399	LOW	Lowes Companies Inc.	1.882
61671	VLO	Valero Energy Corp.	0.854
62092	TMO	Thermo Electron Corp.	1.120
62498	WST	West Inc.	0.305
63976	GMP	Green Mountain Pwr Corp.	0.118
64020	LUB	Lubys Cafeterias Inc.	0.521
64450	NJR	New Jersey Res.	0.326
66325	SLM	Student Loan Marketing Assn.	5.327
66587	HUG	Hughes Supply Inc.	0.065
67213	EKR	E Q K Realty Investors 1	0.026
67248	GMI	Gemini Ii Inc.	0.159
67491	EP	Enserch Exploration Partners Ltd.	0.882
68451	GMH	General Motors Corp.	2.430
69593	NWS	News Corp. Ltd.	1.067
69809	CHS	Chaus Bernard Inc.	0.070
70885	NW	National Westminster Bank Plc.	0.078
71159	TWN	Taiwan Fund Inc.	0.159
71175	UNM	U N U M Corp.	2.787
71300	AHE	American Health Pptys Inc.	0.428
71810	DNP	Duff & Phelps Utilities Income	1.392
72688	SFB	Standard Federal Bank Troy Mich.	0.554
75168	SWZ	Swiss Helvetia Fund Inc.	0.129

Table 3. Continued.

Perm. number	Ticker symbol	Company name	Market capitalization
75193	UTH	Union Texas Petroleum Holdings I.	1.584
75239	NCA	Nuveen California Mun Value Fd.	0.199
75240	NNY	Nuveen New York Mun Value Fund.	0.125
75254	GVT	Dean Witter Government Income Tr.	0.437
75261	MNI	Mcclatchy Newspapers	0.082
75275	TTF	Thai Fund Inc.	0.234
75279	TFB	Municipal Income Trust Ii	0.294
75331	BKT	Blackrock Income Trust Inc.	0.550
75334	HIS	C I G N A High Income Shs.	0.176
75376	AMF	A C M Managed Income Fd Inc.	0.174
75427	TGG	Templeton Global Govts Income Tr.	0.189
75475	VLT	Van Kampen Mer Ltd Trm Hi Inc. Tr.	0.065
75880	FBR	First Brands Corp.	0.558

entire sample period for the 100 securities individually. We also estimate the market model for each sub-period separately.

We also use the weekly return series to calculate the variance of the individual securities ($\sigma^2_{\eta_j}$) and that of the S&P 500 index (σ^2_θ). Given the underlying model on the fundamentals in Eq. (1), we calculate the variance of the idiosyncratic component as follows:

$$\sigma^2_{\varepsilon_i} = \sigma^2_{\eta_i} - \beta_i^2 \sigma^2_\theta. \tag{17}$$

When used in the regression analysis, we divide the estimated $\sigma^2_{\varepsilon_i}$ by 5 to obtain the daily variance. Note, however, that in the empirical analysis, we use standard deviation rather than the variance of the stock.

To obtain the level of liquidity trading in each stock we could have attempted to use total trading volume as a proxy for it. However, observed trading volume of a stock depends on both strategic trading and as non-strategic trading. Strategic trading itself depends on other independent variables in our model for example β and σ^2_ε. Consequently, instead of using trading volume we choose to use market capitalization as a proxy for the level of liquidity trading in a stock.[24] The rationale is based on the observation that large firms attract more non-strategic trading, for example from index and pension funds, than similar but small firms. Higher levels of non-informed trading in large firms can also be motivated by the relatively low transaction costs in these firms.

[24] We are extremely grateful to the referee for pointing us in this direction.

Table 1 reports the summary statistics for the above mentioned variables. The average of the 9900 cross autocorrelation (CAC) individual estimates over the entire sample period is 2.64%. When the analysis is repeated for each year separately the average CAC ranges from 1.21% (in 1992) to 4.29% (in 1990). All the estimates are significant at the 1% level. Note that there is significant variability amongst the individual estimates of cross autocorrelation. The estimates range from a minimum of -25.68% and a maximum of 36.49%. The standard deviation for the entire sample period is 3.49%. Similarly the average of the 100 individual estimates of β over the entire sample period is 0.814 and the sub-period averages range from 0.379 (in 1994) to 0.872 (in 1991). The average size of daily idiosyncratic shock, as measured by σ_{ε}, is 0.018 over the entire sample period. In the sub-periods it ranges from 0.016 to 0.020. Even though the figures are not reported in the table, it is worthwhile to note that the portion of the price innovation attributable to idiosyncratic shocks has been in general increasing over the sample period from 80% (in 1990) to 92% (in 1994). Over the entire sample period, 86% of the price innovation is coming from the idiosyncratic shocks. The randomly selected sample has firms ranging from quite small (market capitalization \$3 million) to quite large (market capitalization \$14.4 billion). The average market capitalization over the entire sample period is \$1.56 billion.

Given the above choices for the independent variables we are in position to run our cross sectional regression. The regression model we use is:

$$CAC_{ij} = \alpha + \gamma_1\beta_i + \gamma_2\beta_j + \gamma_3\sigma_{\varepsilon_i} + \gamma_4\sigma_{\varepsilon_j} + \gamma_5\, Mkcap_i + \gamma_6\, Mkcap_j \qquad (18)$$

In the model CAC_{ij} is the one day lagged cross autocorrelation between daily returns when stock i leads stock j. Specifically, it is the correlation of the stock i's return at time t with that of stock j at time $t+1$. β_i is the market model coefficient of the leading stock with respect to the S&P 500 index, while β_j is the β coefficient of the lagging stock. σ_{ε_i} is the relative weight of the idiosyncratic shocks for the leading stock, while σ_{ε_j} is that for the lagging stock. Finally, $Mkcap_i$ is the market capitalization of the leading stock, while $Mkcap_j$ is that of the lagging stock. In our estimation we normalize the $Mkcap$ such that its unit is \$10 billion. Market capitalization, as discussed earlier, is used to proxy the level of liquidity trading in individual stocks. The theoretical model predicts γ_1 and γ_5 to be positive, and γ_3 and γ_6 to be negative. Since there are 9900 CACs, there are 9900 observations per sample period.

Table 2 documents the results from the regression analysis. The last column reports the results of the cross sectional regression for the entire sample period. The other columns present results of the independent estimation for the five sub-periods. The F-values of the model are significant at the 1% level for the entire sample period and also for all five sub-periods.

The first thing we would like to point out in Table 2 is that our results are consistent with the lead–lag effect identified by the extant literature. Over the entire sample period the coefficient on market capitalization of the leading stock (γ_5) is positive while that on the market capitalization of the following stock (γ_6) is negative. Both coefficients are significant at the 1% level. When we analyze the individual sub-periods the signs of γ_5 and γ_6 remain unchanged. However, due to increased noise we lose some statistical power. In the sub-period analysis γ_5 is significant at the 1% level in three out of five cases while γ_6 is significant in only two. These findings are consistent with the predictions of the model.

Given the above estimates on γ_5 and γ_6, all else being equal, CAC_{ij} is predicted to be higher than CAC_{ji} if the market capitalization of stock i is greater than that of stock j. Although this empirical regularity is well known, we believe that we have provided a better theoretical underpinning for the size based lead–lag effect. Thus, if large stocks are indeed proxying for stocks which have a higher amount of liquidity trading then our theoretical result provides an explanation as to why large stocks might lead small stocks. In the previous section of the paper, we argue that an increase in σ_u^2 increases the level of common factor trading in the stock. An increase in the level of common factor trading in the leading stock increases the CAC while that in the following stock reduces the CAC. A caveat is in order here. This implication is based on our assumption that the amount of liquidity trading is correlated with the size of the firm, which in our opinion is a reasonable assumption.

Interestingly, it appears that cross autocorrelation is more sensitive to the firm size of the lagging stock than that of the leading stock. For the whole sample period, γ_5 is 0.007 whereas γ_6 is -0.016. This implies that if the market capitalization of a leading stock increases by \$10 billion, the cross autocorrelation would increase by 0.7%, whereas if the same size increase occurs to the lagging stock's market capitalization, the cross autocorrelation would decrease by 1.6%.

The prediction of model that the level of cross autocorrelation is positively related to the β of the leading stock is strongly supported by the data. The coefficient of β_i, γ_1, is positive and statistically significant at the 1% level for the entire sample period and also for each of the five sub-periods. Thus, the larger the β of the leading stock, the larger the expected cross autocorrelation. The data, thus, seem to be consistent with the premise of this paper that stocks with a higher β_i have a higher amount of common factor trading, which increases the amount of common factor information impounded in the price and, consequently, increases the CAC_{ij}. The parameter estimate ranges from 0.008 to 0.052 for the sub-periods and is 0.028 for the entire sample period. We would like to note that this result is obtained after controlling for the well-known lead–lag effect due to firm size. Thus, β cannot be proxying for firm size in this regression.

The results also strongly support the prediction that the level of cross autocorrelation is negatively related to σ_ε of the leading stock. The coefficient of σ_{ε_i}, γ_3, is negative and statistically significant at the 1% level not only for the entire sample period but also for each sub-period. The estimates range from

-0.364 to -1.026 for the individual sub-periods, and $\gamma_3 = -0.705$ for the entire sample period. Thus, the data supports the intuition that an increase in the weight of the idiosyncratic component in stock i decreases the level of common factor trading in that stock and consequently reduces CAC_{ij}. A 1% (0.01) increase in the σ_ε of the leading stock will bring about 0.7% drop in the cross-autocorrelation.

As previously mentioned, the theory predicts that the directional relationships of cross autocorrelation with the lagging stock's β coefficient and the weight of the idiosyncratic shocks is parameter dependent and ambiguous. Controlling for the other variables we use the empirical model to suggest the direction of these relationship in the current dataset. It turns out that cross autocorrelation is positively related to β_j, and negatively related to σ_{ε_j}. The estimate for β_j for the entire sample period is 0.02, which is significant at the 1% level. The estimate is positive for all sub-periods and is significant at the 1% level for four of the five sub-periods. We have argued in the previous section that an increase in β_j has two effects. First, it increases the dependence of the security return on the common factor, which implies that the lagging period price adjustment is more dependent on the common factor. This effect will tend to increase CAC_{ij}. Second, an increase in β_j increases the number of common-factor traders in security j, which increases the amount of information impounded into price in the earlier period. The second effect reduces CAC_{ij}. The empirical results show that the first effect dominates the second effect for the particular dataset.

The estimate of σ_{ε_j} is negative (-0.571) and strongly significant for the entire sample period. When the analysis is repeated over the sub-periods the estimate is negative and significant at the 1% level for four out of the five sub-periods. Similar to the case of β_j an increase in σ_{ε_j} has two effects – (i) it reduces the weight of the common factor; and (ii) it reduces the number of common factor traders in stock j. All else being equal, the first effect will decrease the CAC_{ij} and the second effect will increase it. The empirical results suggest that the first effect is stronger than the second effect.

The results of the regression analysis presented in this section have helped us achieve two broad goals. First and foremost, *all* the directional predictions of the theoretical model are shown to be consistent in this dataset. Specifically, we have confirmed that the cross autocorrelation amongst two stocks is (i) increasing in the β coefficient of the leading stock; (ii) increasing in the intensity of liquidity trading of the leading stock (proxied by the market capitalization); (iii) decreasing in the weight of the idiosyncratic shock of the leading stock; and (iv) decreasing in the intensity of liquidity trading of the following stock. Second, for the variables for which the model predictions were ambiguous, the empirical analysis has shown that the underlying fundamental correlation seems to be more important than the relative increase or decrease in the number of common factor informed traders.

6. Conclusion

In this paper, we have taken a simple but well-known structure and have introduced two different types of traders, the common factor informed and the idiosyncratic component informed. We have shown that the tension among these two groups of informed traders depends on the dependence of the security's fundamentals on the common factor. An exogenous specification of the number of traders misses the tension between the two groups and, as such, is an incomplete picture.

We have studied the effects of explicitly modeling this tension and have shown that the competition between these two groups will make the prices of securities, whose fundamentals have intermediate levels of β, less informative, even though they have the most number of informed traders. We have also identified conditions under which the correlations of security prices will be quite close to the correlations of the underlying fundamental values, in spite of the noise due to liquidity trading.

More importantly, we have shown that some stylized facts about observed cross autocorrelations in short horizon security returns, that cannot be explained by extant models, can arise due to the strategic interaction of these different groups of informed traders. Specifically, we have shown that asymmetric cross autocorrelation can arise due to the strategic entry decisions of informed investors. Stocks with a higher level of liquidity trading, a higher β and a lower idiosyncratic variance are more likely to have a higher level of cross autocorrelations. The predictions of the model are strongly supported by the data.

Although we have been able to explain some empirical patterns by modeling the entry decision of the informed traders, our cost structure is somewhat ad hoc. The question whether there exists a more natural cost structure which limits the factor informed traders to a few securities is left for future research. We believe that the modeling of transaction costs as partially variable costs is an interesting issue, which is also left to be studied in the future.

Appendix A

Proof of Lemma 1. Taking derivatives of $\{\pi_F^i, \pi_i\}_{i=1,\dots,M}$ given in Eqs. (5) and (6) with respect to $\{K_i, N_i, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i}\}_{i=1,\dots,M}$, we can obtain the comparative statics collected in Lemma 1. □

Proof of Lemma 2. From Lemma 1, we have $\partial\pi_F^i/\partial K_i < 0$, $\partial\pi_F^i/\partial N_i > 0$ and $\partial\pi_F^i/\partial\beta_i > 0$. Therefore, for a given σ_{ε_i}, we can find $\bar{\beta}$ such that $\pi_F^i(K_i, \infty, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i}) < T_c$ holds for all $\beta_i < \bar{\beta}(\sigma_{\varepsilon_i})$, $K_i \geqslant 1$, and the informed traders with signal $\tilde{\theta}^t$ do not trade the securities with $\beta_i < \bar{\beta}(\sigma_\varepsilon)$. Similarly, for a given β, there exists a $\bar{\sigma}_\varepsilon(\beta)$ such that $\pi_i(\infty, N_i, \beta, \sigma_{\varepsilon_i}, \sigma_{u_i}) < T_c$ holds for all $\sigma_{\varepsilon_i} < \bar{\sigma}_\varepsilon(\beta)$, $N_i \geqslant 1$, and the informed traders with idiosyncratic signal do not trade the securities with $\sigma_{\varepsilon_i} < \bar{\sigma}_\varepsilon(\beta)$. For the securities that satisfy $\beta_i \geqslant \bar{\beta}(\sigma_{\varepsilon_i})$ and $\sigma_{\varepsilon_i} \geqslant \bar{\sigma}_\varepsilon(\beta_i)$, both types of informed traders trade, and the equilibrium numbers of traders are determined from the solutions of the simultaneous equations of $\pi_F^i(K_i^*, N_i^*, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i}) = T_c$ and $\pi^i(K_i^*, N_i^*, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i}) = T_c$. Since

$$|A| \equiv \begin{vmatrix} \dfrac{\partial \pi_F^i}{\partial K_i} & \dfrac{\partial \pi_F^i}{\partial N_i} \\ \dfrac{\partial \pi^i}{\partial K_i} & \dfrac{\partial \pi^i}{\partial N_i} \end{vmatrix}$$

$$= \frac{\beta_i^2\sigma_\theta^2\sigma_{\varepsilon_i}^2((N_i^*+1)^2(3K_i^*+1)\beta_i^2\sigma_\theta^2 + (K_i^*+1)^2(3N_i^*+1)\sigma_{\varepsilon_i}^2)}{(N_i^*+1)(K_i^*+1)(N_i^*(K_i^*+1)^2\sigma_{\varepsilon_i}^2 + K_i^*(N_i^*+1)^2\beta_i^2\sigma_\theta^2)} > 0$$

always holds, we always have unique solutions of (K_i^*, N_i^*). □

Proof of Proposition 1. From $\pi_F^i(K_i^*, N_i^*, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i}) = T_c$ and $\pi_i(K_i^*, N_i^*, \beta_i, \sigma_{\varepsilon_i}, \sigma_{u_i}) = T_c$, comparative statics can be derived from following set of equations, and Lemmas 1 and 2.

$$A\begin{pmatrix} \dfrac{\partial K_i^*}{\partial(\beta_i^2\sigma_\theta^2)} \\ \dfrac{\partial N_i^*}{\partial(\beta_i^2\sigma_\theta^2)} \end{pmatrix} = -\begin{pmatrix} \dfrac{\partial \pi_F^i}{\partial(\beta_i^2\sigma_\theta^2)} \\ \dfrac{\partial \pi_i}{\partial(\beta_i^2\sigma_\theta^2)} \end{pmatrix} \Rightarrow$$

$$\frac{\partial K_i^*}{\partial(\beta_i^2\sigma_\theta^2)} = \frac{1}{2}\,\frac{(K_i^*+1)(2K_i^*(N_i^*+1)^2\beta_i^2\sigma_\theta^2 + (K_i^*+1)^2(3N_i^*+1)\sigma_{\varepsilon_i}^2)}{\beta_i^2\sigma_\theta^2((N_i^*+1)^2(3K_i^*+1)\beta_i^2\sigma_\theta^2 + (K_i^*+1)^2(3N_i^*+1)\sigma_{\varepsilon_i}^2)} > 0$$

$$\frac{\partial N_i^*}{\partial(\beta_i^2\sigma_\theta^2)} = -\frac{1}{2}\,\frac{(K_i^*+1)(N_i^*+1)^3}{(N_i^*+1)^2(3K_i^*+1)\beta_i^2\sigma_\theta^2 + (K_i^*+1)^2(3N_i^*+1)\sigma_{\varepsilon_i}^2} < 0$$

$$A\begin{pmatrix} \dfrac{\partial K_i^*}{\partial \sigma_{\varepsilon_i}^2} \\ \dfrac{\partial N_i^*}{\partial \sigma_{\varepsilon_i}^2} \end{pmatrix} = -\begin{pmatrix} \dfrac{\partial \pi_F^i}{\partial \sigma_{\varepsilon_i}^2} \\ \dfrac{\partial \pi_i}{\partial \sigma_{\varepsilon_i}^2} \end{pmatrix} \Rightarrow$$

$$\frac{\partial K_i^*}{\partial \sigma_{\varepsilon_i}^2} = -\frac{1}{2}\,\frac{(N_i^*+1)(K_i^*+1)^3}{(N_i^*+1)^2(3K_i^*+1)\beta_i^2\sigma_\theta^2 + (K_i^*+1)^2(3N_i^*+1)\sigma_{\varepsilon_i}^2} < 0$$

$$\frac{\partial N_i^*}{\partial \sigma_{\varepsilon_i}^2} = \frac{1}{2}\,\frac{(N_i^*+1)(2N_i^*(K_i^*+1)^2\sigma_{\varepsilon_i}^2 + (N_i^*+1)^2(3K_i^*+1)\beta_i^2\sigma_\theta^2)}{\sigma_{\varepsilon_i}^2((N_i^*+1)^2(3K_i^*+1)\beta_i^2\sigma_\theta^2 + (K_i^*+1)^2(3N_i^*+1)\sigma_{\varepsilon_i}^2)} > 0$$

$$A\begin{pmatrix}\dfrac{\partial K_i^*}{\partial T_c}\\ \dfrac{\partial N_i^*}{\partial T_c}\end{pmatrix}=\begin{pmatrix}1\\1\end{pmatrix}\ \Rightarrow$$

$$\frac{\partial K_i^*}{\partial T_c}=\frac{\dfrac{\partial \pi_i}{\partial N_i}-\dfrac{\partial \pi_i}{\partial K_i}}{|A|}<0$$

$$\frac{\partial N_i^*}{\partial T_c}=\frac{\dfrac{\partial \pi_F^i}{\partial K_i}-\dfrac{\partial \pi_F^i}{\partial N_i}}{|A|}<0$$

$$A\begin{pmatrix}\dfrac{\partial K_i^*}{\partial \sigma_{u_i}}\\ \dfrac{\partial N_i^*}{\partial \sigma_{u_i}}\end{pmatrix}=-\begin{pmatrix}\dfrac{\partial \pi_F^i}{\partial \sigma_{u_i}}\\ \dfrac{\partial \pi_i}{\partial \sigma_{u_i}}\end{pmatrix}\ \Rightarrow$$

$$\frac{\partial K_i^*}{\partial \sigma_{u_i}}=\frac{T_c}{\sigma_{u_i}}\frac{-\dfrac{\partial \pi_i}{\partial N_i}+\dfrac{\partial \pi_i}{\partial K_i}}{|A|}>0$$

$$\frac{\partial N_i^*}{\partial \sigma_{u_i}}=\frac{T_c}{\sigma_{u_i}}\frac{-\dfrac{\partial \pi_F^i}{\partial K_i}+\dfrac{\partial \pi_F^i}{\partial N_i}}{|A|}>0$$

From Eqs. (5) and (6), we can see that $\partial\pi_i^F/\partial\sigma_{u_i}=\pi_i^F/\sigma_{u_i}$ and $\partial\pi_i/\partial\sigma_{u_i}=\pi_i/\sigma_{u_i}$ hold, and from the equilibrium conditions of $\pi_F^i=T_c=\pi_i$, the last two equalities are obtained. □

Proof of Proposition 2. Between two normally distributed random variables x and r, we have $\mathrm{Var}(x\,|\,r)=\sigma_x^2(1-Corr(x,r)^2)$. From Eq. (4), we obtain

$$\lambda_i\tilde{y}_i^t=\frac{K_i\beta_i\tilde{\theta}^t}{(K_i+1)}+\frac{N_i\tilde{\varepsilon}_i^t}{(N_i+1)}+\lambda_i\tilde{u}_i^t.$$

Thus, we have

$$\begin{aligned}\mathrm{Var}(\Delta\tilde{\eta}_i^t\,|\,\lambda_i\tilde{y}_i^t)&=\sigma_{\eta_i}^2(1-Corr(\Delta\tilde{\eta}_i^t,\lambda_i\tilde{y}_i^t)^2)\\&=\sigma_{\eta_i}^2\left(1-\left(\frac{\frac{K_i^*}{K_i^*+1}\beta_i^2\sigma_\theta^2+\frac{N_i^*}{N_i^*+1}\sigma_{\varepsilon_i}^2}{\sigma_{\eta_i}\sqrt{\frac{K_i^*}{K_i^*+1}\beta_i^2\sigma_\theta^2+\frac{N_i^*}{N_i^*+1}\sigma_{\varepsilon_i}^2}}\right)^2\right)\\&=\sigma_{\eta_i}^2\left(1-\left(\frac{K_i^*}{K_i^*+1}\frac{\beta_i^2\sigma_\theta^2}{\beta_i^2\sigma_\theta^2+\sigma_{\varepsilon_i}^2}+\frac{N_i^*}{N_i^*+1}\frac{\sigma_{\varepsilon_i}^2}{\beta_i^2\sigma_\theta^2+\sigma_{\varepsilon_i}^2}\right)\right).\end{aligned}$$

Second equality is obtained from

$$\mathrm{Var}(\lambda_i \tilde{y}_i^t) = \left(\frac{K_i^*}{K_i^*+1}\right)^2 \beta_i^2 \sigma_\theta^2 + \left(\frac{N_i^*}{N_i^*+1}\right)^2 \sigma_{\varepsilon_i}^2 + \lambda_i^2 \sigma_{u_i}^2$$
$$= \left(\frac{K_i^*}{K_i^*+1}\right)^2 \beta_i^2 \sigma_\theta^2 + \left(\frac{N_i^*}{N_i^*+1}\right)^2 \sigma_{\varepsilon_i}^2$$
$$+ \frac{K_i^*}{(K_i^*+1)^2} \beta_i^2 \sigma_\theta^2 + \frac{N_i^*}{(N_i^*+1)^2} \sigma_{\varepsilon_i}^2.$$

If we restrict our attention to the set of securities with identical σ_η, expected trading profits earned by informed traders can be rewritten in the following equations in terms of β_i and σ_{η_i} with $\sigma_{\varepsilon_i}^2 = \sigma_\eta^2 - \beta_i^2 \sigma_\theta^2$.

$$\pi_F^i(K_i, N_i, \beta_i, \sigma_\eta, \sigma_{u_i}) = \frac{\beta_i^2 \sigma_\theta^2}{(K+1)^2} \frac{\sigma_{u_i}}{\sqrt{\frac{K}{(K+1)^2}\beta_i^2\sigma_\theta^2 + \frac{N}{(N+1)^2}(\sigma_\eta^2 - \beta_i^2\sigma_\theta^2)}}, \quad (A.1)$$

$$\pi_i(K_i, N_i, \beta_i, \sigma_\eta, \sigma_{u_i}) = \frac{(\sigma_\eta^2 - \beta_i^2 \sigma_\theta^2)}{(N+1)^2} \frac{\sigma_{u_i}}{\sqrt{\frac{K}{(K+1)^2}\beta_i^2\sigma_\theta^2 + \frac{N}{(N+1)^2}(\sigma_\eta^2 - \beta_i^2\sigma_\theta^2)}}. \quad (A.2)$$

Given trading costs and σ_η, we can safely say that the equilibrium number of informed traders trading a security i is uniquely determined by β_i and $K^*(\beta_i)$ and $N^*(\beta_i)$ represent the equilibrium number of traders of each type trading the security of β_i. If trading costs for both types of informed traders are identical, we have the following properties which will be used extensively for other proofs. First, we have

$$K^*(\beta_i) = N^*(\beta_j) \quad \text{and} \quad N^*(\beta_i) = K^*(\beta_j) \quad \text{where} \quad \beta_j^2\sigma_\theta^2 = \sigma_\eta^2 - \beta_i^2\sigma_\theta^2. \quad (A.3)$$

Eq. (A.3) can be obtained from Eqs. (A.1) and (A.2), and equilibrium condition $\pi_F^i(K_i^*, N_i^*, \beta_i, \sigma_\eta) = \pi_i(K_i^*, N_i^*, \beta_i, \sigma_\eta) = T_c$. Second, from the same equilibrium condition, we also obtain

$$K^*(\beta_i) \leqslant (\geqslant) N^*(\beta_i) \quad \text{for} \quad \frac{\beta_i^2\sigma_\theta^2}{\sigma_\eta^2} \leqslant (\geqslant) \frac{1}{2},$$

$$\text{i.e., } \beta_i^2\sigma_\theta^2 \leqslant (\geqslant) \sigma_\eta^2 - \beta_i^2\sigma_\theta^2 = \sigma_{\varepsilon_i}^2, \quad (A.4)$$

$$\frac{(K^*+1)^2}{(N^*+1)^2} = \frac{\beta_i^2\sigma_\theta^2}{\sigma_\eta^2 - \beta_i^2\sigma_\theta^2}. \quad (A.5)$$

From $\mathrm{Var}(\tilde{\eta}_i^t \mid \lambda_i y_i^t)$ and Eq. (A.3), among the securities with identical σ_η we have

$$\mathrm{Var}(\tilde{\eta}_i^t \mid \lambda_i y_i^t) = \mathrm{Var}(\tilde{\eta}_j^t \mid \lambda_j y_j^t) \quad \text{if} \quad \beta_j^2 \sigma_\theta^2 = \sigma_\eta^2 - \beta_i^2 \sigma_\theta^2. \tag{A.6}$$

Given σ_η, we obtain following first derivative of $\mathrm{Var}(\tilde{\eta}_i^t \mid \lambda_i y_i^t)$ with respect to $\beta_i^2 \sigma_\theta^2$.

$$\begin{aligned}
&-\frac{K^*}{1+K^*} - \frac{1}{(1+K^*)^2}\frac{\partial K^*}{\partial(\beta_i^2\sigma_\theta^2)}\beta_i^2\sigma_\theta^2 + \frac{N^*}{1+N^*} \\
&-\frac{1}{(1+N^*)^2}\frac{\partial N^*}{\partial \sigma_{\varepsilon_i}^2}(\sigma_\eta^2 - \beta_i^2\sigma_\theta^2) \\
&\quad = \psi[\beta_i^2\sigma_\theta^2(N_i^*+1)^2(4K_i^*N_i^* + 2N_i^* - 3K_i^* + 1) \\
&\qquad - (\sigma_\eta^2 - \beta_i^2\sigma_\theta^2)(K_i^*+1)^2(4K_i^*N_i^* + 2K_i^* - 3N_i^* - 1)] \\
&\quad = \psi\beta_i^2\sigma_\theta^2(N_i^*+1)^2(N_i^* - K_i^*)(3K_i^* + 3N_i^* + 2)
\end{aligned}$$

where

$$\psi = \frac{1}{2(K_i^*+1)(N_i^*+1)((N_i^*+1)^2(3K_i^*+1)\beta_i^2\sigma_\theta^2 + (K_i^*+1)^2(3N_i^*+1)(\sigma_\eta^2 - \beta_i^2\sigma_\theta^2))}.$$

Second equality is derived from Eq. (A.5). We can see from Eq. (A.4) that given σ_η, $\mathrm{Var}(\tilde{\eta}_i^t \mid \lambda_i y_i^t)$ increases in β_i for $\beta_i < \hat{\beta} = \sigma_\theta/\sqrt{2}\sigma_\eta$, is maximized at $\beta_i = \hat{\beta}$ and then decreases in β_i for $\beta_i > \hat{\beta}$.

Given σ_η, if we take derivative of $(K^* + N^*)$ with respect to $\beta^2\sigma_\theta^2$, we have following equation from Proposition 1:

$$\begin{aligned}
&\psi\Bigg[2K_i^*(K_i^*+1)(N_i^*+1)^2 - 2N_i^*(N_i^*+1)(K_i^*+1)^2 \\
&+ \frac{\sigma_\eta^2 - \beta_i^2\sigma_\theta^2}{\beta_i^2\sigma_\theta^2}(K_i^*+1)^3(3N_i^*+1) - \frac{\beta_i^2\sigma_\theta^2}{\sigma_\eta^2 - \beta_i^2\sigma_\theta^2}(N_i^*+1)^3(3K_i^*+1)\Bigg] \\
&= \psi(K_i^*+1)(N_i^*+1)[2(N_i^* - K_i^*) + (N_i^*+1)(3N_i^*+1) \\
&\quad - (K_i^*+1)(3K_i^*+1)].
\end{aligned}$$

The equality is obtained from Eq. (A.5), and we can see from Eq. (A.4) that given σ_η, $K_i^* + N_i^*$ is concave in β_i, and it is maximized at $\beta_i = \hat{\beta}$. □

Proof of Proposition 3. From Eq. (12), we have

$$\mathrm{Cov}(\Delta P_i^t, \Delta P_j^t) = E\left[\left(\frac{K_i^*}{K_i^*+1}\beta_i\tilde{\theta}^t + \frac{1}{K_i^*+1}\beta_i\tilde{\theta}^{t-1}\right)\left(\frac{K_j^*}{K_j^*+1}\beta_j\tilde{\theta}^t + \frac{1}{K_j^*+1}\beta_j\tilde{\theta}^{t-1}\right)\right]$$

$$= \frac{K_i^* K_j^* + 1}{(K_i^*+1)(K_j^*+1)}\beta_i\beta_j\sigma_\theta^2.$$

The first equality is derived from independence among random variables. From $\mathrm{Var}(\Delta P_i^t) = \sigma_{\eta_i}^2$ and

$$Corr(\Delta\tilde{\eta}_i^t, \Delta\tilde{\eta}_j^t) = \frac{\mathrm{Cov}(\beta_i\tilde{\theta}^t + \tilde{\varepsilon}_i^t,\ \beta_j\tilde{\theta}^t + \tilde{\varepsilon}_j^t)}{\sigma_{\eta_i}\sigma_{\eta_j}}$$

$$= \frac{\beta_i\beta_j\sigma_\theta^2}{\sigma_{\eta_i}\sigma_{\eta_j}},$$

$$Corr(\Delta\tilde{\eta}_i^t, \tilde{\theta}^t) = \frac{\mathrm{Cov}(\beta_i\tilde{\theta}^t + \tilde{\varepsilon}_i^t,\ \tilde{\theta}^t)}{\sigma_{\eta_i}\sigma_\theta}$$

$$= \frac{\beta_i\sigma_\theta}{\sigma_{\eta_i}}$$

and $Corr(\Delta P_i^t, \Delta P_j^t)/Corr(\Delta\tilde{\eta}_i^t, \Delta\tilde{\eta}_j^t)$ in the proposition is obtained. Since $\mathrm{Cov}(\Delta P_i^t, \Delta P_j^t)$ increases in K_i^* (K_j^*) and K_i^* (K_j^*) is an increasing function of β_i (β_j), the result follows. □

Proof of Proposition 4. From Eq. (12), we derive

$$\mathrm{Cov}(\Delta P_i^t, \Delta P_j^{t+1}) = \mathrm{E}\left[\left(\frac{K_i^*}{K_i^*+1}\beta_i\tilde{\theta}^t\right)\left(\frac{1}{K_j^*+1}\beta_j\tilde{\theta}^t\right)\right]$$

$$= \frac{K_i^*}{(K_i^*+1)(K_j^*+1)}\beta_i\beta_j\sigma_\theta^2.$$

Cross autocorrelations given in the proposition are obtained from $\mathrm{Var}(\Delta P_i^t) = \sigma_{\eta_i}^2$ and $Corr(\Delta\tilde{\eta}_i^t, \tilde{\theta}^t) = \beta_i\sigma_\theta/\sigma_{\eta_i}$. From Proposition 1, we know that K_i^* is an increasing function of β_i and $\mathrm{Cov}(\Delta P_i^t, \Delta P_j^{t+1})$ increases β_i. Similarly, since K_i^* (K_j^*) increases in σ_{u_i} (σ_{u_j}), $\mathrm{Cov}(\Delta P_i^t, \Delta P_j^{t+1})$ is an increasing (decreasing) function of σ_{u_i} (σ_{u_j}). Combining these results with $\mathrm{Var}(\Delta P_i^t) = \sigma_{\eta_i}^2$ and $\mathrm{Var}(\Delta P_j^t) = \sigma_{\eta_j}^2$, comparative statics in the proposition are obtained. □

References

Admati, A., Pfleiderer, P., 1988. A theory of intraday trading patterns: Volume and trading variability. Review of Financial Studies 1, 3–40.

Bossaerts, P., 1993. Transaction prices when insiders trade portfolios. Working Paper, California Institute of Technology.

Bulow, J., Geanakoplos, J.D., Klemperer, P.D., 1985. Multimarket oligopoly: Strategic substitutes and complements. Journal of Political Economy 93 (3), 488–511.

Chan, K., 1992. A further analysis of the lead–lag relation between the cash market and stock index futures market. Review of Financial Studies 5, 123–152.

Chan, K., 1993. Imperfect information and cross-autocerrelation among stock prices. Journal of Fiannce 48 (4), 1211–1229.

Cho, J., Shin, J., Singh, R., 1998. The generality of spurious predictability. Working paper, Hong Kong University of Science and Technology.

Kyle, A.S., 1985. Continous auctions and insider trading. Econometrica 53, 1315–1335.

Lo, A.W., MacKinlay, A.C., 1990a. An econometric analysis of nonsynchronous trading. Journal of Econometrics 45, 181–211.

Lo, A.W., MacKinlay, A.C., 1990b. When are contrarian profits due to stock market overreaction?. Review of Financial Studies 3, 175–206.

Subrahmanyam, A., 1991a. Risk aversion, market liquidity, and price efficiency. Review of Financial Studies 4, 417–441.

Subrahmanyam, A., 1991b. A theory of trading in stock index futures. Review of Financial Studies 4, 17–52.