Review of Economic Studies (2013) **80**, 35–72 doi:10.1093/restud/rds020

Advance access publication 23 April 2012

Informed Trading and Portfolio Returns

ALEX BOULATOV
University of Houston

and

TERRENCE HENDERSHOTT and DMITRY LIVDAN
Haas School of Business,
University of California, Berkeley

First version received January 2010; *final version accepted June* 2011 (*Eds.*)

We solve a multi-period model of strategic trading with long-lived information in multiple assets with correlated innovations in fundamental values. Market makers in each asset can only condition their pricing functions on trading in each asset. Using daily non-public data from the New York Stock Exchange, we test the model's predictions on the conditional and unconditional lead–lag relations of institutional order flow and returns within portfolios. We find support for the model prediction of positive autocorrelations in portfolio returns as well as the predictions for how informed order flow positively predicts future returns and future informed order flow. We show that these relations strengthen for portfolios formed from assets within the same industry, which likely have higher correlation of fundamental values. Furthermore, we discuss issues that arise when testing implications of strategic models with imperfect proxies for the underlying strategic behaviour.

Key words: Market efficiency, Strategic trading, Information asymmetry, Learning

JEL Codes: D4, G1, G12, G14

1. INTRODUCTION

Asset prices following a random walk is the basis for much of the theoretical and empirical asset pricing in financial economics. Therefore, an examination of the predictability of asset returns is the first non-introductory chapter in Campbell, Lo and Mackinlay (1997). Lo and Mackinlay (1988) show that the autocorrelations of equal-weighted portfolio returns are significantly positive, while the autocorrelations of individual asset returns are generally negative. Together these point to positive cross-autocorrelations in asset returns (Lo and Mackinlay, 1990a).

Multiple explanations have been proposed for the positive autocorrelations in portfolios. We study whether informed institutional trading is a source of these correlation patterns.[1] We construct and test a multi-period (Kyle, 1985) style model of strategic trading with long-lived information in multiple assets with positively correlated fundamental values. The model's key assumption is that in each trading period, prices in each asset are functions of trading only in that asset (and not trading in the other asset).

1. The next section describes the related literature.

The simple friction that market makers can only condition their pricing rule on trading in that asset exists in some form in virtually all markets.[2] It can also arise from any friction that prevents market makers from observing and perfectly interpreting information from trading in all other assets continuously and instantaneously, for example, limited attention or cognition.

The model has an analytical solution in the case of two assets traded over two trading rounds when assets have symmetrically distributed fundamental values, liquidity trades are independent across time, and assets have the same variance. The informed trader minimizes the informational impact of his trades by strategically choosing the trading intensities across both time and assets. The informed trader does this across time by curbing the aggressiveness of his trades in both assets in the first period in a manner similar to a two-period version of the original (Kyle, 1985) model. His cross-asset strategy is quite different. The informed trader sets the sensitivities of his trading demand in either asset to be positive in his information about the asset he trades in and negative in his information about the other asset. This strategy hinders the market maker's ability to learn about the asset values.

The model implies that within asset returns are independent across time, whereas the cross-asset returns, and, therefore, returns on a portfolio of assets, are positively autocorrelated. The individual assets' returns are independent across time because asset-specific order flows are sufficient statistics for prices of their respective assets. The informed trader trades only on his residual information advantage, making current order flow independent of lagged returns.

The following example illustrates the intuition behind the positive cross-asset autocorrelation. In the first trading round the market maker in the first asset receives a large positive order flow in her asset but cannot condition on order flow in the second asset. Because the order flow contains a noisy liquidity component, the market maker adjusts her asset's price only partially upwards in response to the large positive order flow. After the transaction takes place, she examines the transaction price of the second asset and uncovers its order flow that contains additional information about her asset. If it is also large and positive, she is more confident about the favourable information and adjusts her asset's price further upwards. If it is not, she is less confident about the favourable information and revises the asset's price downwards. Therefore, the price change of the first asset in the second trading round is positively correlated with the price change of the second asset in the first trading round.

Transaction prices equal market makers' expectations of the asset value conditional on all information available to them. Hence, it is not possible to construct a profitable trading strategy based on the positive autocorrelation in portfolio returns without an information set superior to the market makers. In this sense, prices are weak-form efficient despite the existence of predictability in portfolio returns. However, if investors could trade on cross-asset information before market makers adjust their prices, then profits based on public information exist.

The model's long-lived information provides a rich set of empirical predictions beyond the positive portfolio return autocorrelations. Except for the zero correlation between informed order flow and lagged returns, the model predicts that portfolio lead–lag autocorrelations between returns and informed order flow increase in the correlation between the innovations in the assets' fundamental values. The model also allows for joint tests of the dynamics of returns and informed order flow in a conventional vector autoregression (VAR) setting.

Testing models of strategic informed trading is difficult because the informed trader hides among the uninformed traders. To test the model's predictions, we use daily non-public data on buy and sell volume by institutions on the New York Stock Exchange (NYSE). The data set contains detailed information on all executed orders for 7 years from January 1999 to December

2. Allowing orders to be contingent on trading in multiple assets can prevent market clearing as the contingencies on the different orders can be mutually exclusive.

of 2005. Initially, we use institutional order flow (buy volume minus sell volume) normalized by the market cap as a proxy for the model's informed order flow (Section 5 extends this). We first test the model's predictions on the market portfolio. We find support for the model's predictions: positive autocorrelations in portfolio returns, informed order flow positively predicts future returns and future informed order flow, and returns do not predict future informed order flow.

We also test and find support for the model's comparative static predictions on the fundamental correlation of assets by examining portfolios formed from stocks in the same industry versus portfolios formed from stocks randomly chosen from different industries. We show that the lead–lag correlations between returns and informed order flow is higher in industry portfolios than in randomly formed portfolios.

The model also generates predictions for the joint dynamics of returns and order flow. We confirm the model's comparative static predictions by estimating VARs on within-industry portfolios and randomly formed portfolios. The relevant VAR coefficients on the within-industry portfolios are higher and lower than the same coefficients for the random portfolios. However, the imperfection of our proxy, that is, not all institutional trading is informed, shows up in some of the coefficients not having the predicted signs.

We utilize two approaches to better align the model and data. First, as suggested by the model, we refine our institutional proxy for informed trading to include aggressive orders rather than both aggressive and passive orders. Estimation using the refined proxy produces VAR coefficients matching those from the theory. Second, we demonstrate the intuition behind how removing the passive institutional trades affects the theoretical predictions. These approaches to aligning the theory and empirical results highlight challenges in testing strategic models with imperfect data.

The paper is organized as follows. Section 2 discusses related literature. Section 3 presents the model. Section 4 reports empirical testing. Section 5 discusses the challenges of testing strategic models with imperfect proxies for strategic behaviour and presents two ways to align the model's implications with the empirical results. Section 6 concludes.

2. RELATED LITERATURE

Given the importance of prices following a random walk in asset pricing, there exists a significant empirical and theoretical literature examining positive autocorrelations in portfolio returns. Explanations other than asymmetric information and the slow diffusion of information have been suggested for the positive autocorrelation in daily and weekly portfolio returns. Conrad and Kaul (1988) propose time-varying risk premia. Lo and Mackinlay (1990b), Boudoukh, Richardson and Whitelaw (1994), and Bernhardt and Davies (2008) explore asynchronous trading. However, these have generally proven unsatisfactory. Anderson *et al.* (2008) use intraday transaction data to show that partial price adjustment is the major source of the portfolio autocorrelations. Lo and Mackinlay (1990b), Brennan, Jagadeesh and Swaminathan (1993), Chan (1993), and others suggest that slow adjustment to common information is the source for partial price adjustment.[3]

On the theoretical side, without modelling informed trading, Chan (1993) shows how imperfect updating/learning of signals by market makers about a common factor in stock returns leads to positive autocorrelations in portfolios. Including informed trading in related models allows for additional and more detailed empirical predictions relating return autocorrelations to private information and trading. Bernhardt and Mahani (2007) use the same friction, our model

3. Badrinath, Kale and Noe (1995) and Sias and Starks (1997) show how the fraction of firms owned by institutions affects portfolio return autocorrelations but do not examine the relations between institutional flow and returns that is the focus of this paper.

employs for imperfections in market makers' learning from all signals—market makers can only condition their pricing function on order flow in each asset—to generate positive autocorrelation in portfolio returns. In Bernhardt and Mahani's model, informed trading occurs each period and information is short lived, as the true value of the asset is announced before the start of the subsequent trading round. Positive cross-autocorrelation arises because the price change due to trading in each asset is related to the new information subsequently announced in the other asset. In Bernhardt and Mahani, this occurs when the true values of the assets are announced, whereas in our model, it is when order flows/prices in the other assets are observed.

The differences between using short-lived and long-lived information manifest themselves in the order flow dynamics. If information is short lived, then informed order flow is independent across time and past returns are unrelated to past informed order flow.[4] If information is long lived, as in our model, then informed trading becomes positively autocorrelated. Without a proxy for informed order flow, this difference is not empirically relevant.

Our institutional order proxy for informed trading is strongly autocorrelated and consistent with information being long lived. Beyond this, in a VAR of returns and informed order flow, our long-lived information model predicts that in the order flow equation, the coefficient on lagged returns should be negative, whereas the coefficient on lagged informed order flow should be positive, consistent with our empirical findings in Table 6.

Our paper extends the single-period multi-asset models, for example, Admati (1985) uses a competitive model and Caballe and Krishnan (1994) solve a strategic model, to long-lived information and multiple periods. Bernhardt and Taub (2008) use features of both Admati and Caballe and Krishnan to analyse a model with multiple informed traders and short-lived information. Proposition 8 of Bernhardt and Taub shows that in a two-asset two-speculator problem, the cross-asset order flows are negatively correlated even though the two assets' fundamental values are positively correlated. Section 3.3 discusses how our model with only one informed trader produces the related result that strategic cross-trading leads to lower, although still positive, cross-asset order flow correlation. We show how in strategic models the informed trader negatively trades on his cross-asset information, for example, if the value of one security is positive, the trader is more likely to sell the other security in order to reduce the cross-asset price impact of his trades.

The empirical testing of multi-asset strategic trading models is limited. One significant challenge is that the optimal informed trading strategy generally leads to publicly observable total order flow (informed plus uninformed) having the same correlation structure across time and across assets as the exogenous uninformed order flow (see, *e.g.* Bernhardt and Taub, 2008). This implies that total order flow is only related to returns/price changes contemporaneously through the price impact functions/matrices. Hence, without data providing a finer partition than total order, such as the data used in this paper, tests of multi-asset strategic trading models can only test the contemporaneous price impact function across assets. Pasquariello and Vega (2011) do this via comparative statics on the price impact function with respect to informational heterogeneity and the number of speculators. The dispersion of analyst earnings forecasts and the number of analysts are used as proxies for informational heterogeneity and the number of speculators.

The predictions for informed order flow are much richer, as shown in Sections 3.4 and 4.2. These require a proxy for informed trading for which we use institutional trading. We study how the imperfections in our proxy complicate matching the theoretical predictions to the empirical results in Section 5.

4. Chordia, Sarkar and Subrahmanyam (2011) examine short-lived information with the assumption that common factor information is only traded on in the large stocks to analyse time variation in liquidity and information diffusion across large and small stocks.

3. THE MODEL

3.1. *Set-up*

We consider an economy in which two risky assets are traded in the financial market over two trading rounds. There are three types of risk-neutral agents in the economy: an informed trader, competitive market makers, and liquidity ("noise") traders.[5] At $t = 0$, the informed trader learns the fundamental values of both risky assets, $\mathbf{V} = (V_1, V_2)^{'}$, simultaneously drawn from the joint normal distribution

$$\widetilde{\mathbf{V}} = \begin{pmatrix} \widetilde{V}_1 \\ \widetilde{V}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_0^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right). \tag{1}$$

The statistical properties of the assets' fundamental values are summarized in Assumption 1.[6]

Assumption 1. *The fundamental values of the assets, $\widetilde{V}_{1,2}$, are positively correlated, $\rho > 0$, have means normalized to zero, and the same variance, σ_0^2.*

Each asset is handled by competitive market makers. Market makers do not know V_1 and V_2 until the announcement after the second and final trading round. Before any trading takes place, market makers know the unconditional joint distribution of $\widetilde{\mathbf{V}}$ and thus quote prices $\mathbf{P}(0) = \mathrm{E}[\widetilde{\mathbf{V}}] = \mathbf{0}$ at $t = 0$. The aggregate inelastic liquidity demand in asset $i = 1, 2$ at time $t = 1, 2$, $\mathbf{u}(t) = (u_1(t), u_2(t))^{'}$, is normally distributed with zero mean and variance equal to σ_u

$$\mathbf{u}(t) \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}), \tag{2}$$

where $\mathbf{I}$ is 2×2 identity matrix. Assumption 2 outlines the cross-sectional and inter-temporal properties of the liquidity demands. This assumption simplifies the market makers' inference problem but is not crucial.

Assumption 2. *The liquidity demands are independent across assets,* $\mathrm{Cov}(u_i(t), u_j(t)) = 0$, *and time,* $\mathrm{Cov}(u_i(1), u_i(2)) = 0$.

Figure 1 illustrates the sequence of events and the information sets of all agents. In the first trading round, the informed trader takes into account the correlation between assets by conditioning his demand in each asset, $\mathbf{x}(t) = (x_1(t), x_2(t))^{'}$, on the value of the other asset, $V_{j \neq i}$. The informed order flows at $t = 1$ are given by

$$\mathbf{x}(1) = \begin{pmatrix} \beta_{11}(1) & \beta_{12}(1) \\ \beta_{21}(1) & \beta_{22}(1) \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \beta(1)\mathbf{V}, \tag{3}$$

and the corresponding total order flows, $\mathbf{y}(t) = (y_1(t), y_2(t))^{'}$, are given by

$$\mathbf{y}(1) = \beta(1)\mathbf{V} + \mathbf{u}(1). \tag{4}$$

5. We assume perfect competition between market makers. See Dennert (1993), Biais, Martimort and Rochet (2000), and others for models incorporating imperfect competition between market makers. See Kyle (1989) and others for models of competing informed traders. See Bernhardt and Taub (2006) for a comparison of the Kyle (1985) and Kyle (1989) models. For a review of related models, see Biais, Glosten and Spatt (2005).

6. This and other assumptions are made for the sake of clarity and tractability. Under these assumptions, the model can be solved analytically. By reducing the degrees of freedom, these assumptions also put a much higher hurdle on the model fit to the data. We discuss the positive correlation across asset values further in Section 4.3. Overall information is positively correlated across stocks, but positive correlation is more difficult to show for private information.

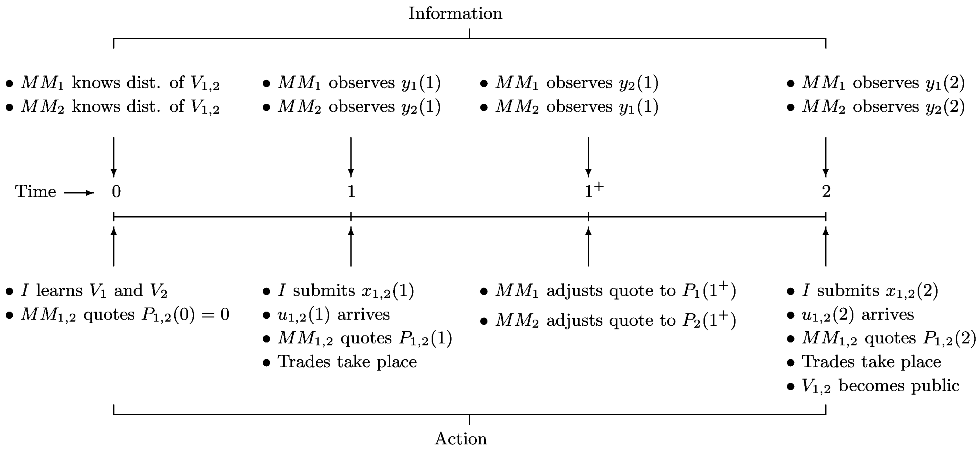

FIGURE 1
This figure shows the timeline of the model

During the trading rounds, market makers observe order flows and use Bayes rule to update their beliefs about V_1 and V_2. Assumption 3 (A3 thereafter) outlines the strategies available to the market makers.

Assumption 3. *In both trading rounds, the market makers condition their pricing functions in an asset on order flow in that asset (and not on order flow in the other asset). After the first trading round, time $t=1^+$, market makers observe the prices of both assets, infer order flow in both assets, and adjust prices before the second trading round.*

At a high enough frequency, A3 is consistent with the fact that no market structure allows market makers (or any traders) to condition their prices in one asset on trading in another asset. Even if market makers could condition their pricing function in an asset on order flow in other assets, A3 is a reduced-form way of capturing any friction in cognition precluding the market makers from instantaneously and fully processing and acting on all information in all securities.[7] The informed trader acts strategically to take advantage of this inefficiency in order to maximize his expected profits.

A3 is crucial for our analysis. If market makers observe and condition prices on order flows in both assets, then, as in Kyle (1985), prices would be fully informationally efficient: information in the current order flow for either asset is orthogonal to the information in the next period order flow for both assets. As a result, without A3, order flows and, therefore, returns for individual assets and their equal-weighted portfolio are not predictable as shown in Remark 1 in Section 3.4.

When A3 holds, market makers condition prices only on their own order flow and prices are less than fully informationally efficient. While each asset price is an unbiased estimate of the true asset value conditional on its own order flow, the sum of asset prices (the price of an equal-weighted portfolio) is not an unbiased estimate of the value conditional on both order flows.

7. Bernhardt and Davies (2008) and Chordia, Sarkar and Subrahmanyam (2011) utilize this same assumption, which is a natural way to produce the imperfect updating/learning in Chan (1993). Biais, Hombert and Weill (2010) and Cespa and Foucault (2010) examine other implications of traders having limited cognition and attention.

The inefficiency arises from information in the cross-asset order flows not being incorporated immediately. Consequently, when market makers update prices using order flows inferred from the price change of the other assets, stock returns may be positively cross-autocorrelated. Implicitly, we assume that no investors can learn from cross-asset order flows before the market makers' adjust their prices.

In accordance with A3, competitive market makers observe the order flow in each asset and set prices at $t=1$ according to the schedules

$$\mathbf{P}(1)=\begin{pmatrix}\mathrm{E}[\widetilde{V}_1|y_1(1)]\\ \mathrm{E}[\widetilde{V}_2|y_2(1)]\end{pmatrix}=\lambda(1)\mathbf{y}(1). \tag{5}$$

In the spirit of Kyle (1985), the inverse market depth parameter $\lambda_i(1)$ in equation (5) is the slope coefficient in the linear regression of $\widetilde{V}_i$ on $y_i(1)$:

$$\begin{aligned}\lambda(1)&=\begin{pmatrix}\frac{\mathrm{Cov}(\widetilde{V}_1,y_1(1))}{\mathrm{Var}(y_1(1))} & 0\\ 0 & \frac{\mathrm{Cov}(\widetilde{V}_2,y_2(1))}{\mathrm{Var}(y_2(1))}\end{pmatrix}\\ &=\begin{pmatrix}\frac{\beta_{11}(1)+\rho\beta_{12}(1)}{\beta_{11}^2(1)+2\rho\beta_{11}(1)\beta_{12}(1)+\beta_{12}^2(1)+\beta_K^2} & 0\\ 0 & \frac{\beta_{22}(1)+\rho\beta_{21}(1)}{\beta_{22}^2(1)+2\rho\beta_{22}(1)\beta_{21}(1)+\beta_{21}^2(1)+\beta_K^2}\end{pmatrix},\end{aligned} \tag{6}$$

where $\beta_K\equiv\frac{\sigma_u}{\sigma_0}$ refers to the elasticity of informed order flow to private information in Kyle (1985). At the end of the first trading round, $t=1^+$, after trades take place, the market makers observe the prices of all assets. Market makers use these prices to infer first-period order flows and adjust the prices to the full information level according to

$$\mathbf{P}(1^+)=\mathrm{E}[\widetilde{\mathbf{V}}|y_1(1),y_2(1)]=\begin{pmatrix}\Lambda_{11} & \Lambda_{12}\\ \Lambda_{21} & \Lambda_{22}\end{pmatrix}\begin{pmatrix}y_1(1)\\ y_2(1)\end{pmatrix}=\Lambda\mathbf{y}(1), \tag{7}$$

where Λ_{ij} are slope coefficients in the linear regression of $\widetilde{V}_i$ on $y_j(1)$:

$$\Lambda_{i1}=\frac{\mathrm{Cov}(\widetilde{V}_i,y_1(1))\mathrm{Var}(y_2(1))-\mathrm{Cov}(\widetilde{V}_i,y_2(1))\mathrm{Cov}(y_1(1),y_2(1))}{\mathrm{Var}(y_1(1))\mathrm{Var}(y_2(1))-(\mathrm{Cov}(y_1(1),y_2(1)))^2}, \tag{8}$$

$$\Lambda_{i2}=\frac{\mathrm{Cov}(\widetilde{V}_i,y_2(1))\mathrm{Var}(y_1(1))-\mathrm{Cov}(\widetilde{V}_i,y_1(1))\mathrm{Cov}(y_1(1),y_2(1))}{\mathrm{Var}(y_1(1))\mathrm{Var}(y_2(1))-(\mathrm{Cov}(y_1(1),y_2(1)))^2},\quad i=1,2.$$

Evaluating the slope coefficients yields

$$\Lambda=\Psi^{-1}\beta(1)', \tag{9}$$

where the covariance matrix of the fundamental values conditional on the order flows, $\mathrm{Cov}\left(\widetilde{\mathbf{V}}|\,\mathbf{y}(1)\right)$, is

$$\Psi=\begin{pmatrix}\psi_{11} & \psi_{12}\\ \psi_{21} & \psi_{22}\end{pmatrix}=\begin{pmatrix}\frac{\beta_K^2}{1-\rho^2}+\beta_1(1)\beta_1'(1) & -\frac{\rho\beta_K^2}{1-\rho^2}+\beta_1(1)\beta_2'(1)\\ -\frac{\rho\beta_K^2}{1-\rho^2}+\beta_1(1)\beta_2'(1) & \frac{\beta_K^2}{1-\rho^2}+\beta_2(1)\beta_2'(1)\end{pmatrix}, \tag{10}$$

with $\beta_i(1)$ denoting the ith row of the matrix $\beta(1)$. The "efficient" prices (equation (7)) are the basis for quoted prices at the start of the second trading round.

At the beginning of the second trading round, $t=2$, the informed trader submits market orders

$$\mathbf{x}(2)=\beta(2)(\mathbf{V}-\mathbf{P}(1^{+})), \tag{11}$$

and market makers receive the following total order flows:

$$\mathbf{y}(2)=\beta(2)(\mathbf{V}-\mathbf{P}(1^{+}))+\mathbf{u}(2). \tag{12}$$

As in the first trading period, market makers can only condition their pricing functions on the order flow in each asset and not on the order flow in the other asset:

$$\mathbf{P}(2)=\mathbf{P}(1^{+})+\lambda(2)\mathbf{y}(2), \tag{13}$$

where the inverse market depth parameter $\lambda(2)$ in equation (13) is the slope coefficient in the linear regression of : $\widetilde{V}_i-P_1(1^{+})$ on $y_i(2)$:

$$\lambda(2)=\begin{pmatrix}\frac{\mathrm{Cov}(\widetilde{V}_1-P_1(1^{+}),y_1(2))}{\mathrm{Var}(y_1(2))} & 0\\ 0 & \frac{\mathrm{Cov}(\widetilde{V}_2-P_2(1^{+}),y_2(2))}{\mathrm{Var}(y_2(2))}\end{pmatrix}. \tag{14}$$

At the end of the second trading round, $\mathbf{V}$ is revealed to all agents and pay-offs are realized.

3.2. *Solution*

The model is solved by finding a vector of trading strategies, $\mathbf{x}^*(t)$, which maximizes the informed trader's expected profits over both trading rounds

$$\mathbf{x}^*(t)=\underset{\mathbf{x}(t)}{\arg\max}\ \underbrace{\mathrm{E}\left[\sum_{t=1}^{2}\sum_{i=1}^{2}x_i(t)(V_i-P_i(t))|\widetilde{\mathbf{V}}=\mathbf{V}\right]}_{\text{Informed trader's total expected pay-off}}. \tag{15}$$

The following theorem summarizes the solution of the model. The symmetry of the asset values' statistical properties together with the independence of liquidity demand across assets and time allow for the model to be solved up to a system of non-linear equations.

Theorem 1. *There exists a linear solution to the informed trader's profit maximization problem* (15) *characterized by the following parameters.*

At time $t=1$, *the informed trader's strategy,* $\mathbf{x}(1)=\beta(1)\mathbf{V}$, *is characterized by* $\beta_{11}(1)=\beta_{22}(1)\equiv\beta_+$ *and* $\beta_{12}(1)=\beta_{21}(1)\equiv\beta_-$ *that are the solutions to the system of non-linear equations:*

$$\beta_++\beta_-=\beta_K\left(2\beta_K\lambda_1+\left(2\lambda_2\beta_K(1+\rho)\left(\frac{1}{(1+\rho)}+\left(\frac{\beta_++\beta_-}{\beta_K}\right)^2\right)^2\right)^{-1}\right)^{-1},$$

$$\beta_+-\beta_-=\beta_K\left(2\beta_K\lambda_1+\left(2\lambda_2\beta_K(1-\rho)\left(\frac{1}{(1-\rho)}+\left(\frac{\beta_+-\beta_-}{\beta_K}\right)^2\right)^2\right)^{-1}\right)^{-1}. \tag{16}$$

The market maker pricing function is $\mathbf{P}(1)=\lambda(1)(\beta(1)\mathbf{V}+\mathbf{u}(1))=\lambda(1)\mathbf{y}(1)$, *with*

$$\lambda_1(1)=\lambda_2(1)\equiv\lambda_1=\frac{\beta_+ +\rho\beta_-}{\beta_+^2+2\rho\beta_+\beta_- +\beta_-^2+\beta_K^2}. \tag{17}$$

At time $t=1^+$ *, the asset prices are updated using both order flows* $\mathbf{P}(1^+)=\Lambda\mathbf{y}(1)$ *with*

$$\Lambda_{11}=\Lambda_{22}\equiv\Lambda_+=\frac{\psi_+\beta_+-\psi_-\beta_-}{\psi_+^2-\psi_-^2}, \tag{18}$$
$$\Lambda_{12}=\Lambda_{21}\equiv\Lambda_-=\frac{\psi_+\beta_- -\psi_-\beta_+}{\psi_+^2-\psi_-^2},$$

where $\psi_+=\frac{\beta_K^2}{1-\rho^2}+\beta_+^2+\beta_-^2$ *and* $\psi_-=-\frac{\rho\beta_K^2}{1-\rho^2}+2\beta_+\beta_-$.

At time $t=2$, *the informed trader's strategy is* $\mathbf{x}(2)=\beta(2)(\mathbf{V}-\mathbf{P}(1^+))$ *and the market maker pricing function is* $\mathbf{P}(2)=\mathbf{P}(1^+)+\lambda(2)(\mathbf{x}(2)+\mathbf{u}(2))$, *with*

$$\beta_{11}(2)=\beta_{22}(2)=\frac{1}{\lambda_2}, \tag{19}$$
$$\beta_{12}(2)=\beta_{21}(2)=0,$$

$$\lambda_1(2)=\lambda_2(2)\equiv\lambda_2=\sqrt{\frac{1-\beta_+\Lambda_+-\beta_-\Lambda_- -\rho(\beta_+\Lambda_- +\beta_-\Lambda_+)}{4\beta_K^2}}. \tag{20}$$

The following section studies the properties of the model's solution.

3.3. *Cross-asset strategic trading and market makers' learning*

Figure 2 provides a graphical illustration of the parameters that characterize Theorem 1 as a function of the correlation of the fundamental values of the assets (ρ). The symmetry of the assets and the independence of the liquidity demand across both time and assets leads to a symmetric solution. The friction of the market makers not being able to condition their prices on order flow in both assets leads to a complex first-period trading strategy by the informed trader characterized by equation (16). The informed trader's strategy solves the fixed point problem discussed in the proof that translates into the system of non-linear equations (16). The solution is unique and can be solved numerically. Figure 2 does this by fixing the parameter $\beta_K=1$ and solving equations (16) for all values of ρ.

Panel A of Figure 2 graphs the informed trader's first-period trading strategy. The solution is symmetric and is described by the linear trading intensities in each asset, β_+, and in the other asset, β_-. Panel A has four lines: β_+ and β_- for the model with and without A3. We first focus on the model solution in Theorem 1 with A3. A striking feature of the trading strategy is that the cross-asset trading intensity, β_-, is negative. The intuition for this follows from examining the informed trader's expected first-period profits in asset i, including the impact on second-period profits conditional on his optimal second-period trading strategy ($x_i^*(2)$):

$$x_i(1)(V_i-\lambda_1x_i(1))-P_i(1^+)x_i^*(2),\quad i=1,2. \tag{21}$$

The asset correlation enable the market makers' to infer additional information from trading in both assets. This cross-asset learning causes the informed trader to attempt to better disguise his first-period trading than in a multi-period single-asset model. Because the assets are positively correlated $P_i(1^+)$ is increasing in both $x_1(1)$ and $x_2(1)$. This implies that increasing $x_i(1)$ can increase first-period profits in asset i but decreases second-period profits in both

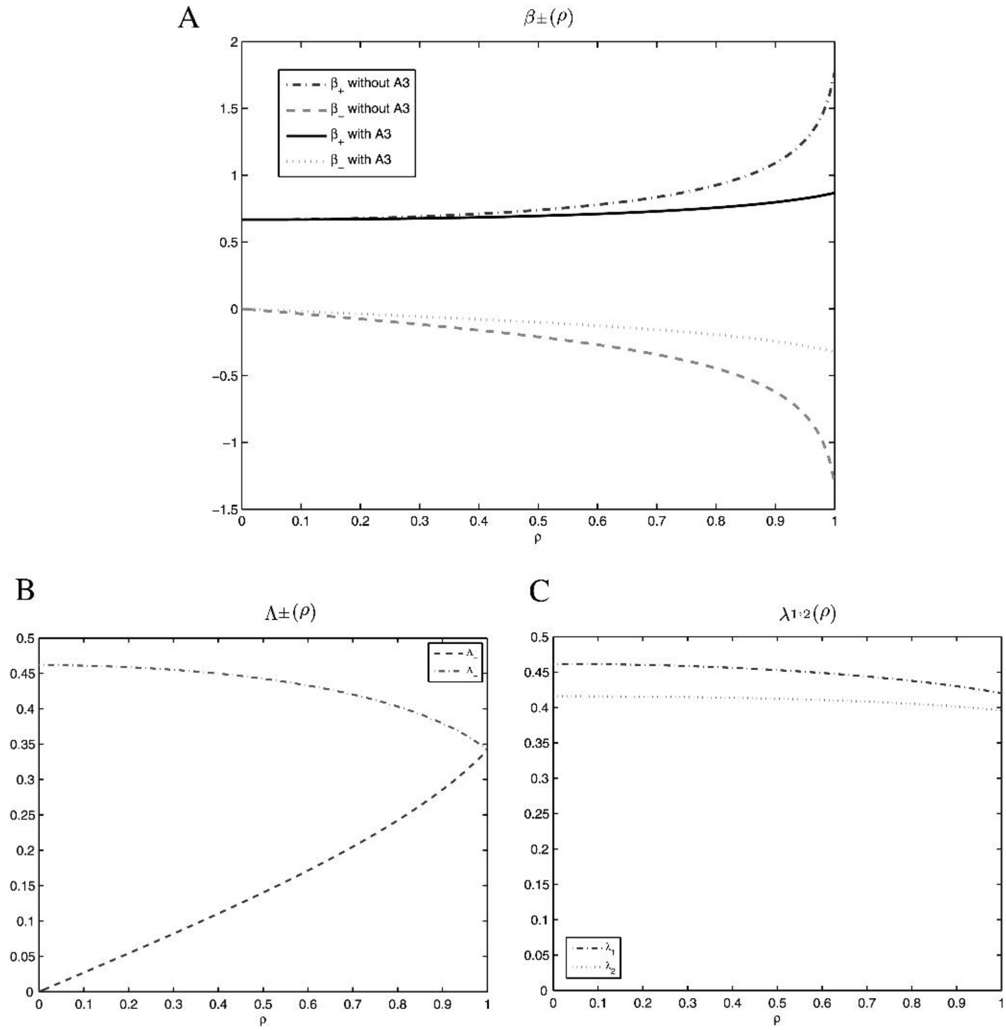

FIGURE 2

We plot the parameters of the solution defined in Theorem 1 as functions of ρ with $\beta_K = 1$. Panel A shows elasticities of the informed order flow to information in the first trading round, $\beta_{\pm}$, with and without A3 given by equations (16) and (A.36), respectively. Panel B shows the within and across asset price impacts of flow, $\Lambda_{\pm}$, given by equation (18). Panel C shows inverse market depth parameter for the two trading rounds: λ_1, given by equation (17), and λ_2, given by equation (20)

assets. This is the basic intuition from a multi-period (Kyle, 1985) model with the addition of the second-period cross-asset price impact. This cross-price impact is lessened by reducing the expected trading intensity is asset i based on the signal about the other asset, that is, by making β_- negative.

When the correlation between assets is zero, β_+ is simply the first-period β in a standard two-period (Kyle, 1985) model and the cross-asset β is zero, $\beta_- = 0$. As the assets become more correlated, ρ increases, the informed trader increases his first-period within-asset trading intensity, β_+. This is because the correlation of the assets improves the market makers' ability to learn when observing prices of both assets. This makes the prices prior to trading in the second-period, $\mathbf{P}(1^+)$, more informative than the first-period transaction prices, $\mathbf{P}(1)$. This decreases the profitability of the informed trader's second-period profits. To compensate for this, the informed trader increases his first-period trading intensity.

The parameters that characterize the market makers' pricing function in Period 1, λ_1 from equation (17), and Period 2, Λ_+, Λ_-, and λ_2 from equations (18) and (20), are graphed in Panels B and C of Figure 2. The optimal trade-off between increasing the within-asset trading intensity β_+ and decreasing the cross-asset trading intensity β_- is reflected in the market makers updating after the first-period trading. The within-asset price impact coefficient is given by Λ_+ and the cross-asset price impact coefficient is Λ_-. At $\rho = 0$, the price impact functions are the standard (Kyle, 1985) two-period solution: $\Lambda_- = 0$ and $\Lambda_+ = \lambda_1$. As the correlation increases, the informed trader tries to hamper the market makers' learning by making β_- more negative.

It is too costly for the informed trader to completely eliminate market makers' learning from cross-asset order flows. This leads to the sensitivity of price to the cross-asset order flows at time $t = 1^+$, Λ_-, increasing in ρ. The informed trader's cross-asset strategy reduces the within-asset price impact, Λ_+. The cross-asset strategy also increases first-period depth as seen by the decrease in the $t = 1$ price impact λ_1. The greater first-period information revelation also leads to the market makers using a smaller second-period price impact λ_2.

Further insight into the informed trader's strategy and how it interacts with the market makers' learning can be gained by examining the model without A3. A3 eliminates market makers' ability to learn about the correlated component in asset values by observing and conditioning their prices on both within- and cross-asset total order flows. This effectively increases the market makers' adverse selection relative to the case without A3. The higher the correlation between assets the more important A3 becomes. In the case of no correlation, $\rho = 0$, A3 is irrelevant.

It is easiest to understand A3's impact by examining a model in which A3 is eliminated in Period 1. The details of the model's solution are in Theorem A2 in the Appendix. Without A3 at time 1, the market makers fully update their prices based on total order flows in both assets, making $\mathbf{P}(1) = \mathbf{P}(1^+)$. The informed trader's expected first-period profits in asset i conditional on his optimal second-period trading strategy, equation (21) can be rewritten as:

$$x_i(1)(V_i - P_i(1)) - P_i(1)x_i^*(2), \quad i = 1, 2. \tag{22}$$

With A3, the expected price in asset i is only a function of order flow in that asset: $P_i(1) = \lambda_1 x_i(1)$. Removing A3 makes prices in both assets at time 1 functions of order flows in both assets. This additional information revelation induces the informed trader to trade more aggressively on his information. Knowing this, the market makers increase the price impact of order flow. This induces the informed trader to increase the intensity of his cross-asset trading. Figure 2 illustrates these effects by graphing the within-asset and across-asset trading intensities, β_+ and β_-, for our model in Theorem 1 and the model without A3 (Theorem A2 in the Appendix).

Removing A3 has significant implications for the dynamics of returns and total order flows. Next we derive those dynamics for the model with A3 and discuss the impact of removing it.

3.4. *Lead–lag relation between order flow and returns*

Using the equilibrium given in Theorem 1, we analyse the lead–lag correlations of individual assets' returns and informed order flows. As is standard, we use incremental price changes, $\Delta P_i(t) = P_i(t) - P_i(t-1)$, as a proxy for assets' returns. Proposition 1 summarizes the within-asset and across-asset lead–lag relations. In the data, we will follow the usual convention of examining autocorrelations and cross-autocorrelations. We present covariances here because the expressions are simpler and have the same sign as the corresponding correlations.

Proposition 1. *The lead–lag covariances of assets' returns are given by*

$$\begin{aligned}
\mathrm{Cov}(\Delta P_i(1), \Delta P_i(2)) &= 0, \quad i=1,2, \\
\mathrm{Cov}(\Delta P_i(1), \Delta P_{j\neq i}(2)) &= \frac{(1-\rho^2)\lambda_1(\lambda_1-\Lambda_+)(\psi_+^2-\psi_-^2)}{\rho(\beta_+^2+\beta_-^2)+2\beta_+\beta_-}\sigma_0^2 > 0, \quad i,j=1,2.
\end{aligned} \tag{23}$$

The lead–lag covariances of informed and total order flows are given by

$$\begin{aligned}
\mathrm{Cov}(x_i(1), x_i(2)) &= \frac{\Lambda_+}{2\lambda_2}\sigma_u^2 > 0,\ i=1,2, \\
\mathrm{Cov}(x_i(1), x_{j\neq i}(2)) &= \frac{\Lambda_-}{2\lambda_2}\sigma_u^2 > 0,\ i,j=1,2, \\
\mathrm{Cov}(y_i(1), y_j(2)) &= 0,\ i,j=1,2.
\end{aligned} \tag{24}$$

The lead-lag covariances of informed and total order flows and returns are given by

$$\begin{aligned}
\mathrm{Cov}(x_i(1), \Delta P_i(2)) &= (2\lambda_1-\Lambda_+)\frac{\sigma_u^2}{2} > 0, \quad i=1,2, \\
\mathrm{Cov}(x_i(1), \Delta P_{j\neq i}(2)) &= -\frac{\Lambda_-}{2}\sigma_u^2+\frac{1}{\lambda_1}\mathrm{Cov}(\Delta P_i(1), \Delta P_{j\neq i}(2)) > 0, \quad i,j=1,2, \\
\mathrm{Cov}(y_i(1), \Delta P_i(2)) &= 0, \quad i=1,2, \\
\mathrm{Cov}(y_i(1), \Delta P_{j\neq i}(2)) &= \frac{1}{\lambda_1}\mathrm{Cov}(\Delta P_i(1), \Delta P_{j\neq i}(2)) > 0, \quad i,j=1,2, \\
\mathrm{Cov}(\Delta P_i(1), x_j(2)) &= \mathrm{Cov}(\Delta P_i(1), y_j(2)) = 0, \quad i,j=1,2.
\end{aligned} \tag{25}$$

The intuition behind the autocovariances and cross-autocovariances is straightforward. Because the informed trader strategically trades so as to not reveal all his information in the first period the within-asset informed order flow positively autocovaries, $\mathrm{Cov}(x_i(1), x_i(2)) > 0$ for $i = 1,2$. Because the assets' final values are positively correlated, the informed order flows have positive contemporaneous covariance and cross-autocovariance, for example, $\mathrm{Cov}(x_1(1), x_2(1)) > 0$ and $\mathrm{Cov}(x_1(1), x_2(2)) > 0$. Because order flow has positive price impact and the uninformed order flow is not observable to the market makers, the informed order flows positively autocovary with subsequent returns: $\mathrm{Cov}(x_i(1), \Delta P_j(2)) > 0$ for $i,j=1,2$.

As is standard in Kyle (1985) type models, asset-specific total order flows within-asset have zero autocovariances. Total order flows are sufficient statistics for return so total order flows and returns in each asset are a martingale: $\mathrm{Cov}(y_i(1), y_i(2)) = \mathrm{Cov}(\Delta P_i(1), \Delta P_i(2)) = 0$ for $i=1,2$. These follow from the informed trader only trading on his residual informational advantage and not on any information already incorporated in prices. This property of the model also makes current informed and total order flows independent of lagged returns: $\mathrm{Cov}(\Delta P_i(1), x_j(2)) = \mathrm{Cov}(\Delta P_i(1), y_j(2)) = 0$ for $i,j=1,2$.

The positive cross-autocovariances in returns and between past total order flows and returns are a consequence of market makers' cross-asset imperfect learning arising from A3. Upon observing the total order flow in one asset, the market makers fully update that asset's price based on the information in order flow. Upon observing the price in the other asset, the market makers can infer the order flow in the other asset. The correlation in the assets fundamental values allows the market makers to use both asset prices/order flows to better filter out the noise trading. If the return in the second asset is consistent with the return in the first asset, then the market makers further update the price in same direction. If the return in the second asset is not consistent with the return in the first asset, then the market makers revise the price in the opposite direction. This leads to positive cross-asset covariances between past order flow and returns and subsequent returns, for example, $\mathrm{Cov}(\Delta P_1(1), \Delta P_2(2)) > 0$ and $\mathrm{Cov}(y_1(1), \Delta P_2(2)) > 0$. The cross-asset covariances are driven solely by the updating as $\mathrm{Cov}(\Delta P_i(1), P_{j\neq i}(1^+) - P_{j\neq i}(1)) > 0$ and $\mathrm{Cov}(\Delta P_i(1), P_i(2) - P_i(1^+)) = 0$ for $i, j = 1, 2$.

A3 drives the delayed price updating based on cross-asset order flows/prices. Without A3, no publicly observable data, for example, total order flows and prices, predict subsequent returns. The following remark summarizes this.

Remark 1. If the market makers can condition their pricing functions on order flows in both assets (removing A3), the following covariances in Proposition 1 become zero (other covariances remain of the same sign):

$$\mathrm{Cov}(\Delta P_i(1), \Delta P_j(2)) = 0, \quad i, j = 1, 2.$$
$$\mathrm{Cov}(y_i(1), \Delta P_j(2)) = 0, \quad i, j = 1, 2,$$

Proposition 1's results on the within- and cross-asset autocovariances provide the intuition necessary to understand the results for a portfolio of the assets. Next we combine both assets into equal-weighted portfolios,

$$\Delta P_p(t) = \frac{1}{2}(\Delta P_1(t) + \Delta P_2(t)), \tag{26}$$
$$x_p(t) = \frac{1}{2}(x_1(t) + x_2(t)),$$

and study their lead–lag covariances in Proposition 2.

Proposition 2. *The equal-weighted portfolio of securities' returns and informed and total order flows have the following lead–lag covariances:*

$$\begin{aligned}\mathrm{Cov}(\Delta P_p(1), \Delta P_p(2)) &= \frac{1}{2}\mathrm{Cov}(\Delta P_1(1), \Delta P_2(2)) > 0,\\ \mathrm{Cov}(x_p(1), x_p(2)) &= \frac{\Lambda_+ + \Lambda_-}{4\lambda_2}\sigma_u^2 > 0,\\ \mathrm{Cov}(y_p(1), y_p(2)) &= 0,\end{aligned} \tag{27}$$

$$\begin{aligned}\mathrm{Cov}(x_p(1), \Delta P_p(2)) &= \left(\lambda_1 - \frac{\Lambda_+ + \Lambda_-}{2}\right)\frac{\sigma_u^2}{2} + \frac{1}{\lambda_1}\mathrm{Cov}(\Delta P_p(1), \Delta P_p(2)) > 0,\\ \mathrm{Cov}(y_p(1), \Delta P_p(2)) &= \frac{1}{\lambda_1}\mathrm{Cov}(\Delta P_p(1), \Delta P_p(2)) > 0,\\ \mathrm{Cov}(\Delta P_p(1), x_p(2)) &= \mathrm{Cov}(\Delta P_p(1), y_p(2)) = 0.\end{aligned}$$

The portfolio autocovariances in Proposition 2 are averages of the within- and across-asset autocovariances in Proposition 1. Past returns in each asset do not covary with subsequent informed or total order flow in either asset so consequently the portfolio returns have zero autocovariance with the portfolio informed and total order flow. Because within-asset returns have zero autocovariance and returns have positive cross-autocovariance, the portfolio returns positively autocovary. Total order flows in each asset predict subsequent informed order flow and returns in both assets so the portfolio informed order flow positively autocovaries with the subsequent portfolio informed order flow and returns. Past portfolio returns and total order flow covary with subsequent portfolio returns because of the imperfect cross-asset price updating due to A3. Without A3, all information in total order flow at time 1 is immediately incorporated into prices at time 1 so past portfolio total order flow and returns do not covary with subsequent portfolio returns.

4. EMPIRICAL TESTS

The model's predicted covariances in Propositions 2 and 3 raise several issues in empirically testing strategic trading models. First, the model is cast in a Kyle (1985) setting where trading occurs in batches in discrete time. A3 prevents efficient cross-asset information incorporation in each trading round. Hence, we need to choose a time horizon corresponding to the trading periods and inefficiency in the model. Based on the daily, weekly, and monthly portfolio return autocorrelations declining as the time horizon increases in Campbell, Lo and Mackinlay (1997), we expect the effects to be largest at shorter horizons. If the positive autocovariance in returns is independent of the time horizon and the variance of returns increases with time horizon, the attenuation of autocorrelations at longer time horizons is consistent with the intuition behind A3.

The positive portfolio return autocorrelations puzzle has typically been studied at a daily or longer horizon.[8] To examine if the imperfect observability/updating friction in A3 is relevant at the horizons previously studied, we analyse daily returns and trading. Below we examine correlations and regressions using lags longer than 1 day. The results are consistent with prior work where the results are stronger at shorter horizons.

A second issue in testing the theory is that neither informed nor total order flows, $x(t)$ and $y(t)$, respectively, are publicly observable.[9] Total order flow and returns only differ by the price impact function. Hence, the lead–lag covariances in Proposition 2 between current portfolio returns and past portfolio returns and past total order flow differ by a constant. As discussed further in Section 4.2, the effective equivalence of total order flow and returns makes their joint dynamics either uninteresting or not well defined. For example, total order flow is unconditionally and conditionally independent of past returns and past total order flow.

To proxy for informed order flow, we will use a measure of institutional trading from the NYSE. Section 5 discusses ways to interpret our proxy as something other than purely informed trading and utilizes an approach to filter institutional trading to yield a better proxy. First utilizing the unfiltered proxy clarifies a number of issues common to testing strategic models.

4.1. *Data*

The data set contains 7 years of daily buy and sell volume of executed institutional investor orders for all common domestic stocks traded on the NYSE between January 1, 1999 and December 31, 2005. The data set was constructed from the NYSE's Consolidated Equity Audit Trail Data (CAUD) files that contain detailed information on all orders that execute on the ex-

8. Anderson *et al.* (2008) and Bernhardt and Davies (2008) use intraday data to examine the impact of asynchronous trading on portfolio return autocorrelations. Anderson *et al.* conduct extensive tests to establish that partial price adjustment, that is, prices not reflecting all cross-asset price information, is the primary cause of positive portfolio return autocorrelations.

9. Papers often use measures of buyer-initiated and seller-initiated trades to proxy for total order flow. Chordia, Roll and Subrahmanyam (2005) examine the within-stock daily and intraday relations between returns and this proxy with a focus on how the relation weakened from 1996 to 2002. Frictions such as short-sale constraints, that is, short sales only being allowed to occur on upticks in price, induce imperfections in proxies for total order flow measured by trade-signing algorithms (Asquith, Oman and Safaya, 2010).

change, both electronic and manual (those handled by floor brokers). One of the fields associated with the buyer and seller of each order, Account Type, specifies whether the order comes from an institutional investor. We use the institutional order flow (buy volume minus sell volume) as a proxy for informed trading. We exclude program trading and index arbitrage trading because in Boehmer and Wu's (2008) individual stock analysis, these do not forecast returns. A sample of the CAUD data was first provided to academics as part of the TORQ data set constructed by Joel Hasbrouck. We complement the CAUD data with daily data on returns (close-to-close returns in data from the Center for Research in Security Prices (CRSP) and closing bid and ask quotes in TAQ), trading volume (CRSP), and market capitalization (number of shares outstanding times price from CRSP).

4.2. *Market-level tests*

We begin testing the lead–lag portfolio relations between informed order flow and returns given in Proposition 2. To do this, we construct an equal-weighted market portfolio for returns and informed order flow for all stocks each day. We use institutional order flow (buy volume minus sell volume) measured as a percent of a total market cap as a proxy for informed order flow in the model. As is standard in the literature, we use daily closing prices to calculate returns.[10] Table 1 provides overview statistics for these market portfolios. The mean market return is 5·7 basis points per day and the standard deviation of the market return is almost 1% per day. Institutions' order flow is slightly positive over the sample period. The standard deviation of institutional order flow is about 1 basis point of market capitalization per day. The last two rows of Table 1 provide evidence on the size of institutional trading volume (as opposed to order flow). Institutional trading is 40·95% of NYSE trading volume. Unless such a large fraction of the total trading is due to informed trading, the institutional trading proxy likely includes some other components as well, for example, hedge funds, investment banks, and broker–dealers providing liquidity. This does not affect the signs of the lead–lag covariances in Proposition 2 but can affect any analysis where both returns and institutional order flow are included as explanatory variables. Section 5 examines how the data and model should relate if institutional order flow is composed of both informed order flow and uninformed liquidity provision.

TABLE 1

This table reports summary statistics of the equal-weighted portfolio of market returns, $R_{\mathrm{m}}(t)$, the equal-weighted portfolio of market institutional order flows (buy volume minus sell volume divided by market capitalization for each stock), $x_{\mathrm{m}}(t)$, as well as ratios of total market institutional volume (buy volume minus sell volume) to market cap and trading volume. $R_{\mathrm{m}}(t)$, $x_{\mathrm{m}}(t)$, and institutional volume are in basis points

	Mean	Std. dev.	Min.	Max.
$R_{\mathrm{m}}(t)$	5·7620	97·3311	−521	517
$x_{\mathrm{m}}(t)$	0·2086	1·0088	−3·9424	5·5299
$\frac{\mathrm{buy}(t)+\mathrm{sell}(t)}{2*\mathrm{mktcap}(t)}$	25·1201	6·8296	5·6341	47·9120
$\frac{\mathrm{buy}(t)+\mathrm{sell}(t)}{2*\mathrm{vol}(t)}$	0·4095	0·0491	0·2704	0·5985

The market portfolios are formed each day by equally weighing returns and normalized order flows for all stocks with available data that day from January 1999 to December 2005. Institutional order flow is measured as a percent of a total market cap.

10. The results are not affected by using the midpoint of the closing bid-ask quotes to calculate returns.

TABLE 2

Panel A of this table reports correlation coefficients of equal-weighted market returns, $R_{\mathrm{m}}(t)$, with lagged equal-weighted market returns and market institutional order flows, $x_{\mathrm{m}}(t)$ and panel B of this table reports correlation coefficients of daily market institutional order flow with lagged equal-weighted market returns and market institutional volume

	$R_{\mathrm{m}}(t-1)$	$R_{\mathrm{m}}(t-2)$	$R_{\mathrm{m}}(t-3)$	$R_{\mathrm{m}}(t-4)$
Panel A: Return				
$R_{\mathrm{m}}(t)$	0·0672	0·0145	0·0456	0·0273
	(2·14**)	(0·43)	(1·34)	(0·79)
$x_{\mathrm{m}}(t)$	−0·0420	−0·0633	−0·0466	−0·0197
	(−0·17)	(−0·21)	(−0·21)	(−0·09)
	$x_{\mathrm{m}}(t-1)$	$x_{\mathrm{m}}(t-2)$	$x_{\mathrm{m}}(t-3)$	$x_{\mathrm{m}}(t-4)$
Panel B: Informed order flow				
$R_{\mathrm{m}}(t)$	0·0326	0·0002	0·0199	0·0269
	(14·40***)	(0·09)	(9·49***)	(11·98***)
$x_{\mathrm{m}}(t)$	0·2548	0·1892	0·2109	0·1827
	(7·81***)	(5·74***)	(7·01***)	(5·82***)

The correlations are reported up to four lags. All variables are daily. Market portfolios are formed each day by equally weighing returns and institutional order flows for all stocks with available data that day from January 1999 to December 2005. Institutional order flow is measured as a percent of a total market cap. Newey–West t-statistics are reported in the parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10% levels, respectively.

We start by testing the lead–lag portfolio correlations from Proposition 2. The model has two trading periods so the calculations in Proposition 2 are for a single lag. As discussed above, it is an open question as to what calendar time period corresponds to the trading periods in the model, so we examine lags up to 4 days. This allows for time- series dependencies in returns and institutional order flow beyond 1 day.

The correlation results are shown in Table 2. Panel A reports the correlation coefficients of equal-weighted market returns, $R_{\mathrm{m}}(t)$, with lagged equal-weighted market returns and market institutional order flow, $x_{\mathrm{m}}(t)$. In agreement with the model's predictions, lagged returns do not predict institutional order flow, but lagged returns predict returns. Panel B reports correlation coefficients of daily market institutional order flow with lagged market returns and market institutional order flow. With the exception of a 2-day lag, correlations between returns and lagged institutional order flow are positive and statistically significant, as the model predicts. Institutional order flow is positively serially correlated, again in agreement with the model.

In the data cross-autocorrelations between informed order flow and returns could be a manifestation of the high autocorrelations of institutional order flow combined with a high contemporaneous correlation between institutional order flow and returns. Under such a scenario, the lagged institutional order flow is a noisy proxy for the lagged returns and, once we control for lagged returns, the lead–lag effect will disappear. To address this alternative hypothesis, we estimate the joint VARs of returns and informed order flow described in Proposition 3.[11]

11. The VAR for returns and total order flow are not informative. The coefficients on returns and total order flow in the total order flow equation are both zero. Proposition 2 shows that contemporaneous returns and contemporaneous total order flow are perfectly collinear with respect to subsequent returns so the coefficients in the return equation are not well-defined. Announcements of public information are one way to reduce the collinearity of returns and total order flow.

Proposition 3. *Consider the following VAR of the equal-weighted portfolios*

$$\begin{pmatrix}\Delta P_p(t)\\ x_p(t)\end{pmatrix}=\begin{pmatrix}a_1\\ a_2\end{pmatrix}+\begin{pmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}\begin{pmatrix}\Delta P_p(t-1)\\ x_p(t-1)\end{pmatrix}+\begin{pmatrix}\varepsilon_1(t)\\ \varepsilon_2(t)\end{pmatrix}. \tag{28}$$

The coefficients b_{ij} are given in the model by

$$\begin{aligned}
b_{11} &= \frac{\Lambda_+ + \Lambda_-}{2\lambda_1} - 1 < 0,\\
b_{12} &= -\left(1+\frac{\beta_K^2}{(1+\rho)(\beta_+ + \beta_-)^2}\right)\lambda_1 b_{11} + 2\beta_K^2\frac{\mathrm{Cov}(\Delta P_1(1), \Delta P_2(2))}{\lambda_1\sigma_u^2(1+\rho)(\beta_+ + \beta_-)^2} > 0,\\
b_{21} &= -\frac{\Lambda_+ + \Lambda_-}{2\lambda_1\lambda_2} < 0,\\
b_{22} &= -\left(1+\frac{\beta_K^2}{(1+\rho)(\beta_+ + \beta_-)^2}\right)\lambda_1 b_{11} > 0.
\end{aligned} \tag{29}$$

The unconditional relations (correlations) in Proposition 2 and the conditional relations (regression coefficients) in Proposition 3 have the same sign when the lagged variable is informed order flow, as b_{12} and b_{22} are positive. However, the conditional relations between past returns and subsequent returns and informed order flow, b_{11} and b_{21}, become negative because of the contemporaneous correlation between the informed order flow and returns. In Proposition 2, there is no relation between past returns and informed order flow. However, liquidity trading is a component of returns. In the second period, the informed trader trades against price changes due to noise. As a result, b_{21} is negative. For the same reason, $\mathrm{Cov}(\Delta P_p(1), \Delta P_p(2))$ is less than $\mathrm{Cov}(x_p(1), \Delta P_p(2))$ that makes b_{11} negative.

Regression models using only returns similar to Proposition 3 are used in Brennan, Jagadeesh and Swaminathan (1993) and Chordia and Swaminathan (2000) to study the lead–lag relations between stocks with different levels of analyst coverage and trading volume. Hou (2007) employs this approach to study lead–lag relations between large and small stocks within and across industries.

Hasbrouck (1991) uses a total order flow proxy from public data together with returns in a VAR at the transaction-level time scale. Hasbrouck measures both the contemporaneous and the lagged price impact of trades. Because our motivation is examining portfolio autocorrelations, we focus on longer horizons and solely on the lagged responses of returns and order flow.[12]

We estimate equation (28) with one ($K=1$) and four (Panel B, $K=4$) lags in the data

$$R_{\mathrm{m}}(t) = b_{0,1} + \sum_{k=1}^{K} b_{11}^k R_{\mathrm{m}}(t-k) + \sum_{k=1}^{K} b_{12}^k x_{\mathrm{m}}(t-k) + \varepsilon_1(t), \tag{30}$$

$$x_{\mathrm{m}}(t) = b_{0,2} + \sum_{k=1}^{K} b_{21}^k R_{\mathrm{m}}(t-k) + \sum_{k=1}^{K} b_{22}^k x_{\mathrm{m}}(t-k) + \varepsilon_2(t). \tag{31}$$

Table 3 reports the VAR results with Newey–West corrected t- and F-statistics. Consistent with the model, the coefficients on institutional order flow in the returns equation, the sum of b_{12}^k, are positive and statistically significant at the 1% level for one-lag regressions (Panel A, 5·1869, t-statistic = 4·72) and at the 5% level for four-lag regressions (Panel B, 7·3214, F-statistic = 3·99). Thus, lagged institutional order flow contains information about future market returns

12. Chordia, Sarkar and Subrahmanyam (2011) estimate the return equation of the VAR using an estimate of total order flow. For the most recent part of their sample, 2001–2007, they do not find order flow to be statistically significant.

TABLE 3

This table reports results of jointly estimating the following one-lag (Panel A, $K=1$) and four-lag (Panel B, $K=4$) VARs, using daily equal-weighted market returns, $R_{\mathrm{m}}(t)$, and institutional order flow, $x_{\mathrm{m}}(t)$, from January 1999 to December 2005

LHS	$R_{\mathrm{m}}(t-1)$	$x_{\mathrm{m}}(t-1)$
Panel A: One-lag		
$R_{\mathrm{m}}(t)$	0·0811	5·1869
	(10·82***)	(4·72***)
$x_{\mathrm{m}}(t)$	0·0003	0·2625
	(1·14)	(119·51***)
LHS	$R_{\mathrm{m}}(t-1:t-4)$	$x_{\mathrm{m}}(t-1:t-4)$
Panel B: Four-lag		
$R_{\mathrm{m}}(t)$	0·1586	7·3214
	*11·97****	*3·99***
$x_{\mathrm{m}}(t)$	−0·0001	0·5062
	0·06	*197·43****

$$R_{\mathrm{m}}(t) = b_{0,1} + \sum_{k=1}^{K} b_{11}^{k} R_{\mathrm{m}}(t-k) + \sum_{k=1}^{K} b_{12}^{k} x_{\mathrm{m}}(t-k) + \varepsilon_1(t),$$

$$x_{\mathrm{m}}(t) = b_{0,2} + \sum_{k=1}^{K} b_{21}^{k} R_{\mathrm{m}}(t-k) + \sum_{k=1}^{K} b_{22}^{k} x_{\mathrm{m}}(t-k) + \varepsilon_2(t).$$

$R_{\mathrm{m}}(t-1:t-k)$, $k=1$ or 4, reports $\sum_{k=1}^{K} b_{11}^{k}$ or $\sum_{k=1}^{K} b_{21}^{k}$, depending on the left-hand side variable. Similarly, $x_{\mathrm{m}}(t-1:t-k)$, $k=1$ or 4, reports $\sum_{k=1}^{K} b_{12}^{k}$ or $\sum_{k=1}^{K} b_{22}^{k}$, depending on the left-hand side variable. All variables are daily. Market portfolios are formed each day by equally weighing returns and institutional order flows for all stocks with available data that day. Institutional order flow is measured as a percent of a total market cap. Newey–West t-statistics are reported in the parentheses. Italics indicate the F-statistics for the hypothesis that the sum of the coefficients equals zero. ***, **, and * denote statistical significance at the 1, 5, and 10% levels, respectively.

beyond that contained in lagged market returns. Also consistent with the model, the coefficients on institutional order flow in the institutional order flow equation, the sum of b_{22}^{k}, are positive and statistically significant at the 1% level, 0·2625 for one lag (Panel A) and 0·5062 for four lags (Panel B). These further support the use of institutional order flow as a proxy for informed order flow and also support the model.

The signs of the coefficients on returns in the returns equation, the sum of b_{11}^{k}, are negative in Proposition 3 while positive and statistically significant in the data at the 1% level for the one-lag (Panel A, 0·0811, t-statistic = 10·82) and the four-lag (Panel B, 0·1586, F-statistic = 11·97) regressions. The coefficients on returns in the institutional order flow equation, the sum of b_{21}^{k}, is negative in Proposition 3. In the VAR, there is no statistically significant relation between returns and subsequent informed order flow. As we will discuss in Section 5, these both may be due to institutional order flow being an imperfect proxy for informed order flow.

4.3. *Comparative statics with respect to asset correlation (ρ): industry tests*

The correlation in the fundamental values of the assets plays an important role in the magnitude of the conditional and unconditional relations between returns and order flows in Propositions 2 and 3. To further study the model's relation to the data, we examine the comparative statics of these coefficients with respect to changing the correlation (ρ). This requires identifying assets that have more and less correlated private information among them.

A common factor beyond the market factor in returns for stocks within an industry is evidence of information being positively correlated across stocks (see King, 1966 for one of the first studies of an industry factor). For ρ to be positive, some portion of that common information must be private.[13] Analysts, hedge funds, and other market participants often specialize in particular industries. Private information production by these participants would raise ρ for stocks within the same industry. Throughout this section, we assume that industry classification identifies stocks with more positively correlated information.

Panel A of Figure 3 presents the correlation coefficients as a function of ρ. As noted in Proposition 2 the covariance of informed order flow with lagged price changes, $\mathrm{Cov}(\Delta P_p(1), x_p(2))$, is zero for all values of ρ. The other correlation coefficients all increase with ρ. These follow from the informed trader's order flow becoming more correlated across assets as the fundamental correlation increases. The informed trader attempting to hinder the market makers' inference problem leads to the autocorrelation of informed order flow growing more slowly than the other two correlations.

Panel B of Figure 3 graphs the VAR coefficients from Proposition 3 as a function of ρ. As seen in Panel A the autocorrelation of the informed trader's trading increases in ρ, which leads to the coefficients on lagged order flow in both the returns and the order flow equations, b_{12} and b_{22}, increasing with ρ. The coefficient on lagged returns in the returns equation, b_{11}, starts off negative at $\rho = 0$ and increases with ρ; b_{11} increases faster than b_{12} and b_{22} because the informed trader increases the negative cross-asset trading with ρ but not fast enough to counteract the market makers' improved cross-asset learning. The coefficient on lagged returns in the informed-order-flow equation, b_{21}, becomes increasingly negative with ρ. This is because the increased correlation of fundamentals means that the informed trader trades more aggressively against noise in the portfolio returns.

To classify stocks, we use the 12 industry SIC-code-based classifications from Ken French's Web site. The final of the 12 industries includes stocks that do not fit the first 11 industries and is referred to as "other." The other group is by far the largest group (763 stocks in our sample period as compared to the next largest which is "Manufacturing" with 314 stocks). To ensure that all our tests do not reuse any data we will use this other category to form random (non-industry) portfolios.

To construct portfolios, we take stocks in the first 11 industry portfolios and calculate returns and informed order flows as done for the market portfolio. For the other category, we randomly divide stocks in it into 11 portfolios and calculate returns and informed order flows for each. Thus, we have 22 portfolios, 11 of which are formed from stocks within the same industry and 11 of which are formed from stocks chosen from random industries. To focus on the industry-level effects, we remove the market-level effects from the within-industry and random portfolios by using the residuals from the regression of returns (order flows) of each portfolio on the market returns (order flows). To avoid creating correlation in the residuals across the industry and random categories, portfolios in each group are regressed on the "market" formed from stocks only within the 11 portfolios in the same category.

Similar to Table 2 for the market-level analysis, Table 4 examines the lead–lag correlations coefficients for the 11 portfolios formed within industries and the 11 randomly formed portfolios. The correlation coefficients for each of the 22 portfolios are estimated individually. To allow for statistical inference that properly accounts for possible correlations across portfolios and time, we estimate the full covariance matrix for the 22 coefficients via seemingly

13. Industry-specific common shocks to supply and demand for firms, or different sensitivities across industries to market-wide shocks, lead to an industry factor in stock returns. Shocks that affect some firms within an industry positively and others negatively would not.

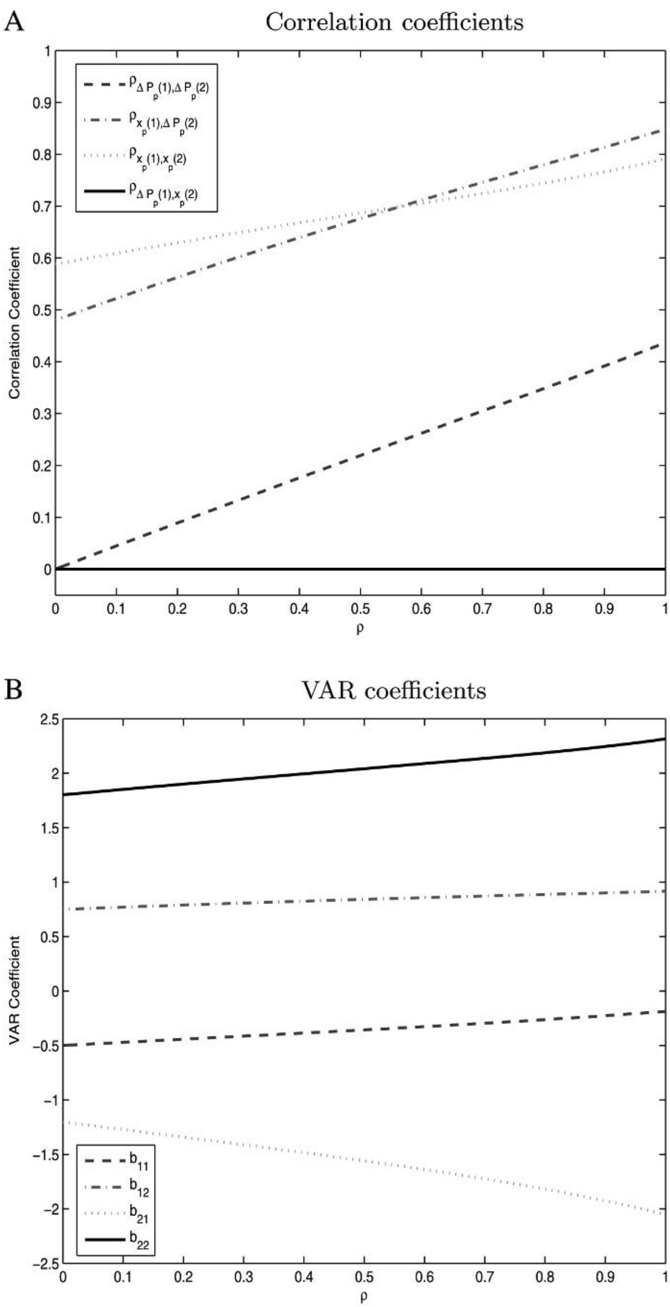

FIGURE 3

This figure provides comparative statics with respect to ρ for the equal-weighted portfolios implied by the model. Panel A plots the various lead–lag correlation coefficients implied by the model as functions of ρ. Panel B plots VAR coefficients from the following regression:

$$\begin{pmatrix} \Delta P_p(t) \\ x_p(t) \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} \Delta P_p(t-1) \\ x_p(t-1) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{pmatrix}$$

as functions of ρ with $\beta_K = 1$

TABLE 4

This table compares lead–lag correlations between portfolios of assets with high and low correlation in fundamental values, ρ

	$R_i(t-1)$	$R_i(t-2)$	$R_i(t-3)$	$R_i(t-4)$	$x_i(t-1)$	$x_i(t-2)$	$x_i(t-3)$	$x_i(t-4)$
Panel A. Within-industry (high ρ)								
$R_i(t)$	0·1066	0·0201	0·017	−0·0003	0·0347	0·0023	0·0150	0·0062
	*222·01******	*7·81******	*5·69******	*0·00*	*24·99******	*0·01*	*3·85****	*0·62*
$x_i(t)$	−0·0103	0·0053	0·02237	0·0158	0·2563	0·1731	0·1482	0·1260
	1·25	*0·64*	*5·56******	*3·50****	*1358·50******	*596·61******	*433·12******	*311·44****
Panel B. Random (low ρ)								
$R_i(t)$	−0·0331	−0·0117	−0·0165	−0·0063	−0·0081	−0·0246	−0·0070	−0·0101
	*21·13******	*2·64**	*5·27******	*0·77*	*1·12*	*12·46******	*0·71*	*1·43*
$x_i(t)$	0·0160	0·0177	0·0188	0·0115	0·2297	0·1270	0·1064	0·0967
	*5·30******	*6·77******	*7·80******	*3·13****	*1079·70******	*317·58******	*220·83******	*182·66******
Panel C. High minus low								
$R_i(t)$	0·1396	0·0318	0·0337	0·0060	0·0427	0·0269	0·0219	0·0163
	*189·48******	*9·76******	*10·95******	*0·35*	*25·16******	*2·28****	*4·56******	*1·44*
$x_i(t)$	−0·0264	−0·0124	0·0048	0·0043	0·0266	0·0461	0·0418	0·0293
	*6·56******	*3·91******	*2·12**	*0·57*	*7·28******	*21·04******	*17·13******	*8·39******

We start with 12 Fama and French industry portfolios using daily data from January 1999 to December 2005. To construct portfolios, we take stocks in the the first 11 industry portfolios and calculate returns and informed order flows as we did for the market portfolio. For the other category, we randomly divide stocks in it into 11 portfolios and calculate returns and informed order flows for each. Thus, we have 22 portfolios, 11 formed from stocks within the same industry and 11 formed from stocks chosen from random industries. To focus on the industry-level effects, we remove the market-level effects from the within-industry and random portfolios by using the residuals from the regression of returns (order flows) of each portfolio on the market returns (order flows). To avoid creating correlation in the residuals across the industry and random categories, portfolios in each group are regressed on the "market" formed from stocks only within the 11 portfolios in the same category. Panel A reports average correlation coefficients of returns, $R_i(t)$, and institutional order flow, $x_i(t)$, with their lagged counterparts for high ρ portfolios. Panel B reports average correlation coefficients of returns, $R_i(t)$, and institutional order flow, $x_i(t)$, with their lagged counterparts for low ρ portfolios. Institutional order flow is measured as a fraction of a total market cap. The correlations are reported up to four lags. Panel C reports the differences between the average correlation coefficients from Panels A and B. *Italics* indicate the F-statistics for the hypothesis that the average correlation coefficient equals zero (Panels A and B) and that the difference is equal to zero (Panel C). Statistical significance is estimated using SUR; ***, **, and * denote statistical significance at the 1, 5, and 10% levels, respectively.

unrelated regression (SUR) with the Newey–West correction to control for heteroscedasticity and autocorrelation.

Panel A of Table 4 reports the correlation coefficients for the 11 within-industry portfolios. Average coefficients across each category of portfolios are reported along with the corresponding F-statistics. As with the market portfolio, returns are significantly positively correlated with lagged returns and lagged informed order flow. Order flow is positively autocorrelated. Order flow is not correlated with the first two lags of returns. These correlations are consistent with the model.

Panel B of Table 4 reports the correlation coefficients for the 11 randomly formed portfolios. Consistent with the model's predictions order flow is positively autocorrelated. Contrary to the model's predictions, order flow is positively and significantly correlated with lagged returns while returns are negatively autocorrelated for the random portfolios. Also, returns and lagged order flows are negatively correlated, although only the one-lag coefficient is statistically significant at 1% level. These results are potentially caused by removing the market-level effects from the random portfolios.

Panel C tests the differences in each correlation coefficient between the industry portfolios (Panel A) and random portfolios (Panel B). The differences in average coefficients between the two categories is given along with the F-statistic that the difference does not equal zero. The differences in the autocorrelation of returns, order flow, and returns and lagged order flow are all positive and statistically significant. These differences match the model predictions that the correlation coefficients are increasing in ρ as shown in Panel A of Figure 4. The difference in the autocorrelation of order flow and lagged returns is negative for first two lags and positive for the next two lags.

Using the same approach as Table 4, Table 5 presents the VAR for returns and order flows for the industry and random portfolios. As in Table 3, the VARs are estimated for both one lag and four lags. As with the correlation coefficients in Table 4, the VAR coefficients for each of the 22 portfolios are estimated simultaneously via SUR.

Panel A of Table 5 presents the one-lag and four-lag VAR for the within-industry portfolios. As with the market portfolio, the coefficients on lagged order flow are positive and statistically significant in both the return and order flow equations. The coefficients on lagged returns in the return equation are positive and statistically significant. Unlike the market results, but consistent with the model's predictions, the coefficients on returns in the order flow equation are negative.

Panel B of Table 5 provides the one-lag and four-lag VAR for the randomly formed portfolios. The cross-equation coefficients, order flow in the return equations and returns in the order flow equations, are both negative but do not differ from zero in the one-lag VAR. As with the correlation coefficients, the coefficients on lagged order flow are positive in the order flow equations and negative in the return equations. Similarly, the coefficients on lagged returns are negative and significant in the returns equation in the one-lag and four-lag specifications.

Panel C tests the differences in the VAR coefficients between the industry portfolios (Panel A) and random portfolios (Panel B). All the differences in coefficients have the signs predicted by the model as shown in Panel B of Figure 4.

5. PROXIES FOR INFORMED TRADING

Overall, the empirical results fit many of the model's predicted signs and comparative statics with respect to ρ for the unconditional and conditional relations between returns and institutional order flow. However, the signs of the coefficients on returns in both the return and order flow equations in the VAR, b_{11} and b_{21}, fit less well. Proposition 3 and Figure 3 show that, conditional on lagged informed order flow, lagged returns should negatively relate to subsequent returns and informed order flow. In the model, all permanent price changes are driven by informed order flow so the returns orthogonal to informed order flow are due to noise trading. The informed trader profits by trading against noise in the subsequent period, causing b_{11} and b_{21} to be negative.

In the empirical work, we proxy for informed trading using institutional trading. Table 1 shows that institutional trading is 40·95% of NYSE trading volume. It is unlikely that all this trading volume represents informed trading. For example, institutions such as hedge funds, investment banks, and broker–dealers also attempt to profit by providing liquidity.

We explore two related ways to better align the model and empirical results. First, we remove the liquidity providing trades from our order flow variable and show how this aggressive institutional order flow measure produces results consistent with the theory.[14] Second, we theoretically show how predictions from the model change if we assume our order flow variable includes some liquidity provision.

14. We are grateful to an anonymous referee for suggesting this.

TABLE 5

This table reports results of jointly estimating the following one-lag (Panel A, $K=1$) and four-lag (Panel B, $K=4$) VARs

	One-lag		Four-lag	
	$R_i(t-1)$	$x_i(t-1)$	$R_i(t-1:t-4)$	$x_i(t-1:t-4)$
Panel A. Within-industry (high ρ)				
$R_i(t)$	0·1028	0·8931	0·1126	0·9883
	*198·81****	*4·91****	*70·93****	*2·47**
$x_i(t)$	−0·0011	0·2685	−0·0011	0·4385
	*58·38****	*1439·68****	*18·39****	*1604·79****
Panel B. Random (low ρ)				
$R_i(t)$	−0·0328	−0·1203	−0·0714	−0·6524
	*20·50***	*0·47*	*22·61****	*5·11****
$x_i(t)$	−0·0003	0·2303	0·0010	0·3599
	0·95	*1073·56****	*2·65**	*992·94****
Panel C. High minus low				
$R_i(t)$	0·1355	1·0134	0·1839	1·6407
	*174·15****	*5·31****	*83·76****	*5·61****
$x_i(t)$	−0·0008	0·0381	−0·0021	0·0786
	*6·41****	*14·63****	*10·33****	*24·69****

We start with 12 Fama and French industry portfolios using daily data from January 1999 to December 2005 which we split into two groups of 11 portfolios each. For within-industry portfolios (high ρ), we use the first 11 industries—everything but the "other" group. Because assets in this case are industry specific, they have a common industry component and thus proxy for assets with high ρ. For the random portfolios (low ρ), we randomly divide stocks in the "other" group into 11 equal-sized portfolios. We then remove a common market component from each portfolio by regressing its return (institutional order flow) on the market return (institutional order flow) and using the residuals. For each portfolio, we calculate the daily equal-weighted returns, $R_i(t)$, and institutional order flow, $x_i(t)$. Institutional order flow is measured as a fraction of a total market cap. We estimate the following VAR for each portfolio:

$$R_i(t) = b_{0,i,1} + \sum_{k=1}^{K} b_{11}^{i,k} R_i(t-k) + \sum_{k=1}^{K} b_{12}^{i,k} x_i(t-k) + \varepsilon_{i,1}(t),$$

$$x_i(t) = b_{0,i,2} + \sum_{k=1}^{K} b_{21}^{i,k} R_i(t-k) + \sum_{k=1}^{K} b_{22}^{i,k} x_i(t-k) + \varepsilon_{i,2}(t).$$

The results in panel A are for portfolios of firms within the same industry. $R_i(t-1:t-k)$, $k=1$ or 4, reports the average $\sum_{k=1}^{K} b_{11}^{i,k}$ or $\sum_{k=1}^{K} b_{21}^{i,k}$, depending on the left-hand side variable. Similarly, $x_i(t-1:t-k)$, $k=1$ or 4, reports the average $\sum_{k=1}^{K} b_{12}^{i,k}$ or $\sum_{k=1}^{K} b_{22}^{i,k}$, depending on the left-hand side variable. *Italics* indicate the F-statistics for the hypothesis that the average of the coefficients equals zero (Panels A and B) and that the difference is equal to zero (Panel C). Coefficients are estimated using SUR; ***, **, and * denote statistical significance at the 1, 5, and 10% levels, respectively.

5.1. *Aggressive institutional order flow as proxy for informed order flow*

Kyle (1985) style models are commonly interpreted as the informed trader using aggressive/ market orders. Therefore, a natural approach to better identifying informed trading is to filter out more passive trades that provide liquidity.

Prior to the NYSE's introduction of widely used fully automated execution at the end of 2006 trades were executed with manual intervention. Orders were sent to floor brokers, typically over the telephone, and sent electronically directly to the market maker, referred to as a specialist on

TABLE 6

This table reports results of jointly estimating the following one-lag (Panel A, $K=1$) and four-lag (Panel B, $K=4$) VARs, using daily equal-weighted market returns, $R_{\mathrm{m}}(t)$, and aggressive institutional order flow, $x_{\mathrm{m}}(t)$, from January 1999 to December 2005

LHS	$R_{\mathrm{m}}(t-1)$	$x_{\mathrm{m}}(t-1)$
Panel A: One-lag		
$R_{\mathrm{m}}(t)$	−0·0956	2·8984
	(2·13**)	(6·93***)
$x_{\mathrm{m}}(t)$	−0·0097	0·3196
	(6·37***)	(24·32***)
LHS	$R_{\mathrm{m}}(t-1:t-4)$	$x_{\mathrm{m}}(t-1:t-4)$
Panel B: Four-lag		
$R_{\mathrm{m}}(t)$	−0·0119	2·4865
	0·01	*1·85***
$x_{\mathrm{m}}(t)$	−0·0196	0·5083
	*8·22****	*22·36****

$$R_{\mathrm{m}}(t)=b_{0,1}+\sum_{k=1}^{K}b_{11}^{k}R_{\mathrm{m}}(t-k)+\sum_{k=1}^{K}b_{12}^{k}x_{\mathrm{m}}(t-k)+\varepsilon_1(t),$$

$$x_{\mathrm{m}}(t)=b_{0,2}+\sum_{k=1}^{K}b_{21}^{k}R_{\mathrm{m}}(t-k)+\sum_{k=1}^{K}b_{22}^{k}x_{\mathrm{m}}(t-k)+\varepsilon_2(t).$$

$R_{\mathrm{m}}(t-1:t-k)$, $k=1$ or 4, reports $\sum_{k=1}^{K}b_{11}^{k}$ or $\sum_{k=1}^{K}b_{21}^{k}$, depending on the left-hand side variable. Similarly, $x_{\mathrm{m}}(t-1:t-k)$, $k=1$ or 4, reports $\sum_{k=1}^{K}b_{12}^{k}$ or $\sum_{k=1}^{K}b_{22}^{k}$, depending on the left-hand side variable. All variables are daily. Market portfolios are formed each day by equally weighing returns and institutional order flows for all stocks with available data that day. Institutional order flow is measured as a percent of a total market cap. Newey–West t-statistics are reported in the parentheses. Italics indicate the F-statistics for the hypothesis that the sum of the coefficients equals zero. ***, **, and * denote statistical significance at the 1, 5, and 10% levels, respectively.

the NYSE. If a number of orders expressing interest in the same security arrived close together, the market maker would hold an informal double auction. Consequently, orders are not marked as demanding or providing liquidity in the NYSE database. Boehmer and Kelley (2009) propose an approach to identify active/aggressive (liquidity demanding) and passive (liquidity supplying) order flow. On page 3584, they write that their "approach exploits the direction of institutional trading relative to the current day's return. Because negative return days arise due to overall selling pressure, institutional buying on those days is likely to provide liquidity and result from a passive trading strategy. Similarly, institutional selling is likely to provide liquidity on positive return days. In contrast, if institutions trade in the same direction as the return, they probably demand liquidity via an aggressive trading strategy." We follow this approach to filter our proxy.

In Section 4, our order flow proxy in each stock each day is institutional buy volume minus sell volume as a percent of total market capitalization. Following the Boehmer and Kelley (2009) approach, we measure aggressive institutional order flow in each stock as buy volume if the stock's return that day is positive, as (minus) sell volume if the stock's return is negative, and zero if the stock's return is zero. Then, as in Section 4, we take the equal-weighted average of these individual stock order flows to construct the market order flow portfolio.

While return volatility remains relatively constant over our 1999–2005 sample period, volatility of the order flow portfolio increases over time. The change in order flow volatility may be due to falling trading costs, increased automation of trading, or other secular trends. To account for this trend, we normalize the market order flow on day t by a rolling estimate of its lagged volatility from days $t-70$ to $t-11$.[15]

Table 6 reports the Table 3 VAR results using the measure of aggressive institutional order flow. The coefficients on returns in both the return and order flow equations in Table 3, b_{11} and b_{21}, respectively, did not align well with the corresponding theoretical predictions in Proposition 3. In Table 3, b_{11} is positive and b_{21} is statistically insignificant. Proposition 3 predicts both should be negative as the component of returns that is orthogonal to contemporaneous informed order flow is noise. This noise in returns is reversed in the subsequent period by the informed trading against it, resulting in b_{11} and b_{21} being negative. In the one-lag VAR in Panel A, both b_{11} and b_{21} coefficients are negative: b_{11} is $-0{\cdot}0956$ (t-statistic = 2·13) and b_{21} is $-0{\cdot}0097$ (t-statistic = 6·37). The coefficients on order flow in both equations are positive, as predicted in Proposition 3 and as in Table 3 using the basic order flow measure. In Panel B with additional lags, all coefficients have the same sign as in Panel A, although b_{11} loses statistical significance.

5.2. *Predictions when institutional trading includes liquidity provision*

We next examine the intuition for how liquidity providing trades in our institutional order flow proxy impacts the model's theoretical predictions.[16] To do this, we keep Assumptions 1–3 and our previous equilibrium results while reinterpreting the measure of institutional order flow using the following assumption.

Assumption 4. *The institutional order flow in each asset, $\widehat{x_i}(t)$, consists of two components: the informed component, $x_i(t)$, and the uninformed liquidity provision component, $l_i(t)$.*

The liquidity provision component corresponds to a fraction $M \in [0,1]$ of trading against total order flow:

$$l_i(t) = -My_i(t) = -M\frac{\Delta P_i(t)}{\lambda_1}, \tag{32}$$

which is liquidity provision at the market price, $P_i(1)$. Therefore, the equal-weighted portfolio of the institutional order flow, $\widehat{x}_p(t)$, equals

$$\widehat{x}_p(t) = x_p(t) - M\frac{\Delta P_p(t)}{\lambda_1}. \tag{33}$$

The appendix calculates the corresponding VAR coefficients for returns, $\widehat{b}_{11}$ and $\widehat{b}_{21}$:

$$\widehat{b}_{11} = (1-M)b_{11} + M\frac{\mathrm{Cov}(x_p(1), \Delta P_p(2))}{\lambda_1 \mathrm{Var}(x_p(1))}, \tag{34}$$

$$\widehat{b}_{21} = (1-M)\left((1-M)b_{21} + M\frac{\mathrm{Cov}(x_p(1), x_p(2))}{\lambda_1 \mathrm{Var}(x_p(1))}\right).$$

Both $\widehat{b}_{11}$ and $\widehat{b}_{21}$ are a combination of the corresponding negative coefficient from the original model, b_{11} and b_{21}, due to informed trading and a positive part related to the liquidity

15. The volatility of the basic order flow measure in Section 4 also exhibits this trend. Applying this same adjustment to the base order flow produces results qualitatively similar to those in Tables 2–5.

16. The model assumes perfect competition among market makers/liquidity suppliers. Relaxing this assumption to allow for imperfect competition is straightforward but tedious and does not qualitatively affect the below results for b_{11} and b_{21}. What is needed for the model to match the empirical signs of b_{11} and b_{21} is a sufficiently large component of the institutional order flow to have a negative contemporaneous correlation with price changes. This can also be accomplished by assuming that the informed trader can condition his trading on contemporaneous noise trading as in Rochet and Vila (1994).

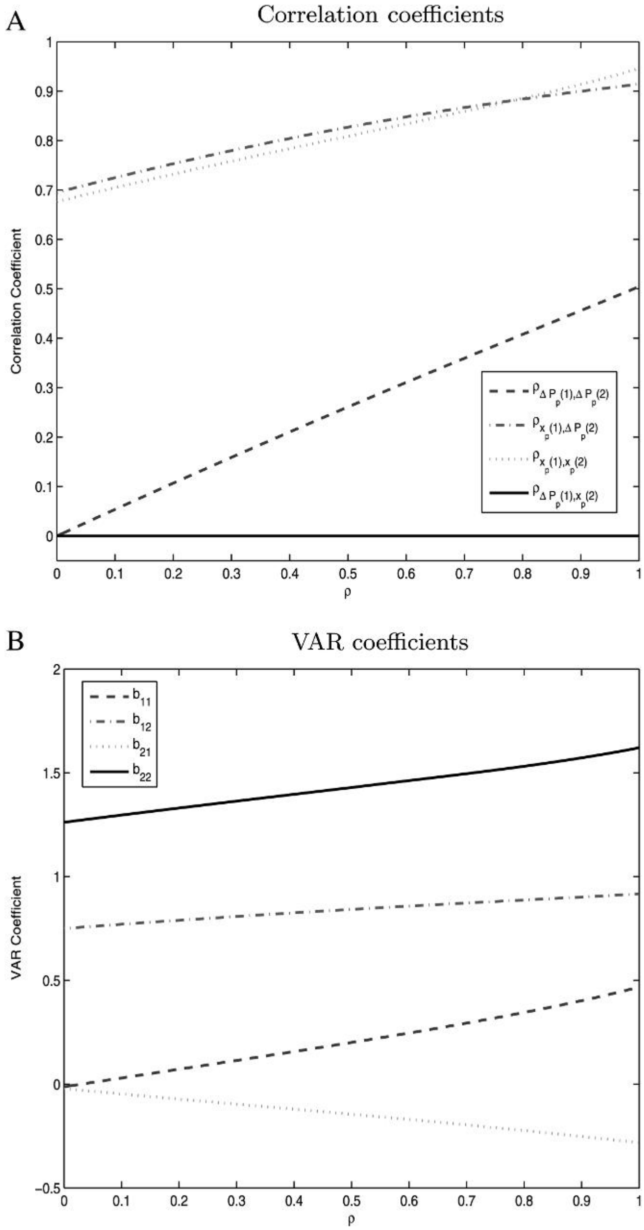

FIGURE 4

This figure provides comparative statics with respect to ρ for the equal-weighted portfolios implied by the extended model with institutional order flow, $\widehat{x}_p(t)$, containing both informed trading and uninformed liquidity provision and defined as

$$\widehat{x}_p(t) = x_p(t) - M\frac{\Delta P_p(t).}{\lambda_1}$$

Panel A plots various lead–lag correlation coefficients implied by the model as functions of ρ. Panel B plots VAR coefficients from the following regression:

$$\begin{pmatrix}\Delta P_p(t)\\ \widehat{x}_p(t)\end{pmatrix} = \begin{pmatrix}a_1\\ a_2\end{pmatrix} + \begin{pmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}\begin{pmatrix}\Delta P_p(t-1)\\ \widehat{x}_p(t-1)\end{pmatrix} + \begin{pmatrix}\varepsilon_1(t)\\ \varepsilon_2(t)\end{pmatrix}$$

as functions of ρ with $\beta_K = 1$ and $M = 0{\cdot}3$

provision component. If the liquidity provision component of institutional order flow is sufficiently large, the VAR coefficient $\widehat{b}_{11}$ and $\widehat{b}_{21}$ change sign and become positive. Figure 4 shows this by graphing the same correlation and VAR coefficients as Figure 3 does for the original model for the extended model with $M = 0{\cdot}3$. The qualitative difference between Figures 3 and 4 is that the intercepts of $\widehat{b}_{11}$ and $\widehat{b}_{21}$ are higher than the intercepts of b_{11} and b_{21}, providing a closer match to the empirical VAR coefficients in Table 5.

This section explored two ways to better align the differences in the theoretical predictions and empirical results that arise when testing models with imperfect measures of strategic behaviour. First, in Table 6, we removed the liquidity providing trades from our order flow variable and showed how the empirical results match the model's predictions. Second, we provided the theoretical intuition for why predictions from the model fit the results in Tables 2–5 if we assume the order flow variable includes some liquidity provision.

6. CONCLUSION

We solve a multi-period model of strategic trading with long-lived information in multiple assets with correlated innovations in fundamental values. The model assumes that market makers cannot condition their price functions in either asset on trading in the other asset. Using daily non-public data from the NYSE, we test the model's predictions on the unconditional and conditional relations between our proxy for informed order flow (institutional order flow) and returns within portfolios. We find support for the model's prediction of positive autocorrelations in portfolio returns as well as the predictions for how informed order flow positively predicts future returns and future informed order flow. We also test the model's comparative static predictions for how the relations between informed order flow and returns depend on the fundamental correlation of assets within a portfolio. We find support for the predicted relations in the data. We discuss and implement related empirical and theoretical approaches to testing strategic models with imperfect proxies.

The predictability of a portfolio of asset returns, but lack of predictability within the individual assets, follows from the market makers' inability to condition their pricing functions in each asset on trading in all the assets. The correlation in the informed order flow across assets reveals additional information which the market makers incorporate into price before the next trading opportunity. Therefore, while returns are predictable, there is no opportunity for anyone without the informed trader's information or without the ability to trade on cross-asset information before market makers adjust their prices to construct a profitable trading strategy based on the positive autocorrelation in portfolio returns. In this sense, our results are consistent with weak-form market efficiency even though there is predictability in portfolio returns.

The automation of the trading process, often referred to as algorithmic trading (Hendershott, Jones and Menkveld, 2011), undoubtedly affects the ability of market makers and investors to process and act on cross-asset information. Reducing the amount of time for the incorporating cross-asset information should reduce the magnitude of portfolio return predictability. Consistent with this for our 1999–2005 sample period, we find the first-order autocorrelation of daily returns on the equal-weighted market portfolio to be $0{\cdot}07$ (Table 6). This is dramatically lower than the autocorrelation of $0{\cdot}35$ for the 1962–1994 time period reported in Campbell, Lo and Mackinlay (1997).

APPENDIX

Proof of Theorem 1. The informed trader's total expected pay-off, π_{I}, is given by

$$\pi_{\mathrm{I}} = \mathrm{E}\left[\sum_{t=1}^{2}\sum_{i=1}^{2} x_i(t)\,(V_i - P_i(t))\,|\widetilde{\mathbf{V}} = \mathbf{V}\right], \tag{A.1}$$

where the expectation is performed with respect to both the liquidity demand and the realizations of the fundamentals.[17]

Combining equations (A.1) with equations (5), (7), and (13), we obtain

$$\begin{aligned}\pi_{\mathrm{I}} &= \mathrm{E}[x_1(1)(V_1-\lambda_1(1)x_1(1))]\\ &\quad+\mathrm{E}[x_1(2)(V_1-(\Lambda_{11}y_1(1)+\Lambda_{12}y_2(1))-\lambda_1(2)x_1(2))]\\ &\quad+\mathrm{E}[x_2(1)(V_2-\lambda_2(1)x_2(1))]\\ &\quad+\mathrm{E}[x_2(2)(V_2-(\Lambda_{21}y_1(1)+\Lambda_{22}y_2(1))-\lambda_2(2)x_2(2))].\end{aligned} \tag{A.2}$$

Following Kyle (1985), the optimization is performed in two steps using backwards induction. First, we optimize equation (A.2) with respect to $x_1(2)$ and $x_2(2)$ to obtain

$$x_i^*(2)=\underset{x_i(2)}{\arg\max}\; x_i(2)(V_i-P_i(1^+)-\lambda_i(2)x_i(2)), \quad i=1,2, \tag{A.3}$$

which after comparing with equation (11) implies that the optimal matrix $\beta(2)$ is diagonal with

$$\begin{aligned}&\beta_{11}(2)=\frac{1}{2\lambda_1(2)}, \quad \beta_{22}(2)=\frac{1}{2\lambda_2(2)},\\ &\beta_{21}(2)=\beta_{12}(2)=0.\end{aligned} \tag{A.4}$$

This is because all agents know that $t=2$ is the last trading round before the fundamentals are revealed. Clearly, this is not the case for the first trading round, and, therefore, we do not expect $\beta(1)$ to be diagonal.

Second, we substitute $x_i^*(2)$ back into equation (A.2) and optimize it with respect to $x_1(1)$ and $x_2(1)$ taking $x_i^*(2)$ as given

$$\begin{aligned}x_1^*(1)&=\underset{x_1(1)}{\arg\max}\, x_1(1)(V_1-\Lambda_{11}x_1^*(2)-\Lambda_{21}x_2^*(2)-\lambda_1(1)x_1(1)),\\ x_2^*(1)&=\underset{x_2(1)}{\arg\max}\, x_2(1)(V_2-\Lambda_{12}x_1^*(2)-\Lambda_{22}x_2^*(2)-\lambda_2(1)x_2(1)).\end{aligned} \tag{A.5}$$

The first-order conditions for $x_i^*(1)$ which follow from equation (A.5) can be written in matrix form as

$$\mathbf{x}^*(1)=\frac{1}{2\lambda(1)}\left(\mathbf{V}-\Lambda'\frac{1}{2\lambda(2)}(\mathbf{V}-\Lambda\mathbf{x}^*(1))\right), \tag{A.6}$$

taking into account that

$$\mathbf{x}^*(2)=\frac{1}{2\lambda(2)}(\mathbf{V}-\Lambda\mathbf{x}^*(1)). \tag{A.7}$$

Combining equation (A.6) with equation (3), $\beta(1)$ can be found as a solution of the following fixed-point condition for

$$\beta(1)=\left(\mathbf{I}-\frac{1}{2\lambda(1)}\Lambda'\frac{1}{2\lambda(2)}\Lambda\right)^{-1}\frac{1}{2\lambda(1)}\left(\mathbf{I}-\Lambda'\frac{1}{2\lambda(2)}\right), \tag{A.8}$$

where $\mathbf{I}$ stands for the unit matrix.

We proceed with a proof in two steps. First, we guess that solution is symmetric

$$\begin{aligned}\lambda_1(1)&=\lambda_2(1)=\lambda_1,\\ \lambda_1(2)&=\lambda_2(2)=\lambda_2,\\ \beta_{11}(1)&=\beta_{22}(1)=\beta_+,\\ \beta_{12}(1)&=\beta_{21}(1)=\beta_-,\\ \Lambda_{11}&=\Lambda_{22}=\Lambda_+,\\ \Lambda_{12}&=\Lambda_{21}=\Lambda_-.\end{aligned} \tag{A.9}$$

Second, we will verify that solution (A.9) exists and satisfies all the necessary conditions.

17. Note that the expectations with respect to $\mathbf{u}(1)$ and $\mathbf{u}(2)$ are "nested" because the liquidity demand in the first trading round is observed in the second trading round and the insider's strategy is conditioned on this.

In the symmetric case, the fixed point condition (A.8) is simplified to

$$\left(\mathbf{I}-\frac{1}{4\lambda_1\lambda_2}\Lambda^2\right)\beta(1)=\frac{1}{2\lambda_1}\left(\mathbf{I}-\frac{1}{2\lambda_2}\Lambda\right), \tag{A.10}$$

and effectively represents a system of two equations for β_+ and β_-:

$$\begin{aligned}\left(1-\frac{\Lambda_+^2+\Lambda_-^2}{4\lambda_1\lambda_2}\right)\beta_+-\frac{\Lambda_+\Lambda_-}{2\lambda_1\lambda_2}\beta_-&=\frac{1}{2\lambda_1}\left(1-\frac{\Lambda_+}{2\lambda_2}\right),\\ -\frac{\Lambda_+\Lambda_-}{2\lambda_1\lambda_2}\beta_++\left(1-\frac{\Lambda_+^2+\Lambda_-^2}{4\lambda_1\lambda_2}\right)\beta_-&=-\frac{\Lambda_-}{4\lambda_1\lambda_2}.\end{aligned} \tag{A.11}$$

Next, we introduce new "auxiliary" variables

$$z_\pm\equiv\sqrt{1\pm\rho}\,\frac{(\beta_+\pm\beta_-)}{\beta_K}. \tag{A.12}$$

The subscript $\pm$ on a variable is used to represent two equations where in the first (second) equation the relevant variable with subscript $\pm$ is replaced by that variable with subscript $+$ $(-)$. Similarly, if $\pm$ is used as an operator, it means $+$ in the first equation and $-$ in the second equation. Therefore, equation (A.12) represents two equations: $z_+=\sqrt{1+\rho}\frac{(\beta_++\beta_-)}{\beta_K}$ and $z_-=\sqrt{1-\rho}\frac{(\beta_+-\beta_-)}{\beta_K}$. z_+ represents a rescaling of the expected informed trading intensity, $\beta_++\beta_-$, and z_- represents a rescaling of the difference between the within and across asset expected informed trading intensity, $\beta_+-\beta_-$. The properties of $z_\pm$ which we will examine later simplify many of the following proofs. Next we guess that

$$\Lambda_+\pm\Lambda_-=\frac{1}{\beta_K}\frac{\sqrt{1\pm\rho}\,z_\pm}{1+z_\pm^2}. \tag{A.13}$$

Substituting both equations (A.12) and (A.13) back into the system (A.11) yields a system of non-linear equations for $z_\pm$. Next we verify that our guess for $\Lambda_\pm$ is self-consistent. It can be easily shown that with the symmetric solution, the matrix Ψ takes the following form:

$$\begin{aligned}\psi_{11}=\psi_{22}\equiv\psi_+&=\frac{\beta_K^2}{2}\left(\frac{1+z_+^2}{1+\rho}+\frac{1+z_-^2}{1-\rho}\right),\\ \psi_{12}=\psi_{21}\equiv\psi_-&=\frac{\beta_K^2}{2}\left(\frac{1+z_+^2}{1+\rho}-\frac{1+z_-^2}{1-\rho}\right).\end{aligned} \tag{A.14}$$

It follows from equation (9) that

$$\begin{aligned}\Lambda_+&=\frac{\psi_+\beta_+-\psi_-\beta_-}{\psi_+^2-\psi_-^2}=\frac{1}{2\beta_K}\left(\frac{\sqrt{1+\rho}\,z_+}{1+z_+^2}+\frac{\sqrt{1-\rho}\,z_-}{1+z_-^2}\right),\\ \Lambda_-&=\frac{\psi_+\beta_--\psi_-\beta_+}{\psi_+^2-\psi_-^2}=\frac{1}{2\beta_K}\left(\frac{\sqrt{1+\rho}\,z_+}{1+z_+^2}-\frac{\sqrt{1-\rho}\,z_-}{1+z_-^2}\right),\end{aligned} \tag{A.15}$$

which verifies equation (A.13). Finally, using equation (A.14), we find λ_1

$$\lambda_1=\frac{\beta_++\rho\beta_-}{\beta_+^2+2\rho\beta_+\beta_-+\beta_-^2+\beta_K^2}=\frac{1}{\beta_K}\frac{\sqrt{1+\rho}\,z_++\sqrt{1-\rho}\,z_-}{2+z_+^2+z_-^2}. \tag{A.16}$$

Using equations (14) and (A.25), we find λ_2

$$\lambda_2^2=\frac{1-\beta_+\Lambda_+-\beta_-\Lambda_--\rho\,(\beta_+\Lambda_-+\beta_-\Lambda_+)}{4\beta_K^2}=\frac{1}{8\beta_K^2}\left(\frac{1+\rho}{1+z_+^2}+\frac{1-\rho}{1+z_-^2}\right), \tag{A.17}$$

which completes the proof. ‖

Lemma A1. $\Lambda_-(\rho)$ *is positive on the interval* $\rho\in(0,1]$.

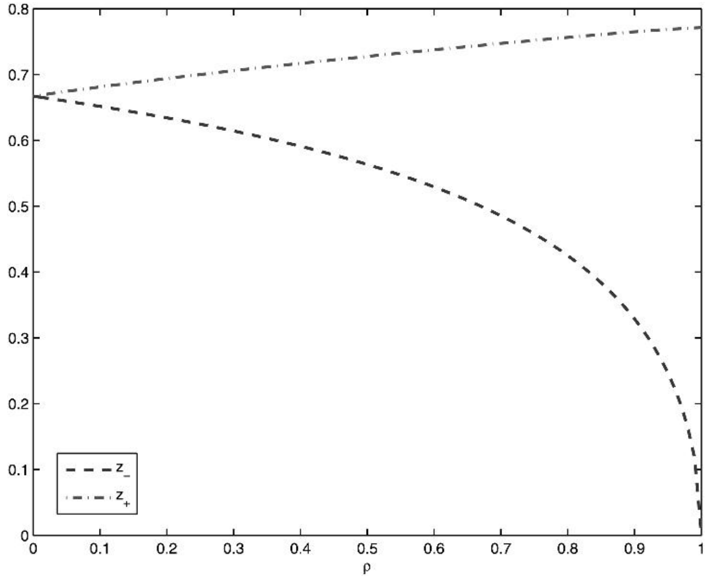

FIGURE A1
Plot of auxiliary variables $z_{\pm}$

Proof of Lemma A1. .

$$\Lambda_{-}(\rho)\Lambda_{+}(\rho) = \frac{1}{4\beta_K^2}\left(\frac{(1+\rho)z_{+}^2}{(1+z_{+}^2)^2} - \frac{(1-\rho)z_{-}^2}{(1+z_{-}^2)^2}\right)$$

$$= \frac{1}{4\beta_K^2}\frac{z_{+}^2(1+z_{-}^2)^2 - z_{-}^2(1+z_{+}^2)^2 + \rho(z_{+}^2(1+z_{-}^2)^2 + z_{-}^2(1+z_{+}^2)^2)}{(1+z_{+}^2)^2(1+z_{-}^2)^2}. \tag{A.18}$$

We need to show that

$$z_{+}^2(1+z_{-}^2)^2 - z_{-}^2(1+z_{+}^2)^2 = (z_{+}-z_{-})(1-z_{+}z_{-})[z_{+}(1+z_{-}^2)+z_{-}(1+z_{+}^2)] > 0, \tag{A.19}$$

which follows from the property described below in equation (A.20).

The below Figure A1 graphs the numerical solution for $z_{\pm}(\rho)$ as functions of ρ. It shows that $z_{+}(\rho)$ is monotonically increasing on the interval $\rho \in (0,1]$, while $z_{-}(\rho)$ is monotonically decreasing on the same interval. Importantly, the auxiliary functions $z_{\pm}(\rho)$ and $z_{+}(\rho)$ satisfy the condition

$$1 > z_{+}(\rho) \geq z_{-}(\rho) > 0, \quad \rho \in (0,1]. \tag{A.20}$$

The result of Lemma A1 follows because $\Lambda_{+}(\rho)$ is positive and $\Lambda_{-}(\rho)\Lambda_{+}(\rho) > 0$ on $\rho \in (0,1]$. ‖

The following Lemma is useful for subsequent proofs.

Lemma A2. *Order flow has the following variances:*

$$\begin{aligned}
&\mathrm{Var}(y_1(1)) = \mathrm{Var}(y_2(1)) = \left(1+\frac{z_{+}^2+z_{-}^2}{2}\right)\sigma_u^2,\\
&\mathrm{Var}(y_1(2)) = \mathrm{Var}(y_2(2)) = 2\sigma_u^2,\\
&\mathrm{Var}(x_1(1)) = \mathrm{Var}(x_2(1)) = \left(\frac{z_{+}^2+z_{-}^2}{2}\right)\sigma_u^2,\\
&\mathrm{Var}(x_1(2)) = \mathrm{Var}(x_2(2)) = \sigma_u^2.
\end{aligned} \tag{A.21}$$

Total order flows, $y_i(t)$, have the following covariances:

$$\begin{aligned}
\mathrm{Cov}(y_1(1),y_2(1)) &= \left(\frac{z_+^2-z_-^2}{2}\right)\sigma_u^2,\\
\mathrm{Cov}(y_1(1),y_1(2)) &= \mathrm{Cov}(y_1(1),y_2(2))=0,\\
\mathrm{Cov}(y_1(2),y_2(2)) &= \mathrm{Cov}(x_1(2),x_2(2))=\frac{\sigma_0^2}{8\lambda_2^2}\left(\frac{1+\rho}{1+z_+^2}+\frac{1-\rho}{1+z_-^2}\right).
\end{aligned} \tag{A.22}$$

Price changes have the following covariances:

$$\begin{aligned}
\mathrm{Var}(\Delta P_1(1)) &= \mathrm{Var}(\Delta P_2(1))=\lambda_1^2\mathrm{Var}(y_1(1)),\\
\mathrm{Var}(\Delta P_1(2)) &= \mathrm{Var}(\Delta P_2(2))=2\lambda_2^2\sigma_u^2+(\Lambda_-^2-(\Lambda_+-\lambda_1)^2)\mathrm{Var}(y_1(1)),\\
\mathrm{Cov}(\Delta P_1(2),\Delta P_2(2)) &= \lambda_2^2\mathrm{Cov}(y_1(2),y_2(2))\\
&\quad+((\Lambda_+-\lambda_1)^2-\Lambda_-^2)\mathrm{Cov}(y_1(1),y_2(1)).
\end{aligned} \tag{A.23}$$

Proof of Lemma A2. $\mathrm{Var}(y_1(1))$ is equal to $\mathrm{Var}(y_2(1))$ by symmetry and can be calculated directly

$$\begin{aligned}
\mathrm{Var}(y_1(1)) &= \sigma_0^2(\beta_+^2+\beta_-^2+2\rho\beta_+\beta_-)+\sigma_u^2\\
&= \frac{\sigma_0^2\beta_K^2}{4}\left\{\frac{2z_+^2}{1+\rho}+\frac{2z_-^2}{1-\rho}+2\rho\left(\frac{z_+^2}{1+\rho}-\frac{z_-^2}{1-\rho}\right)\right\}+\sigma_u^2\\
&= \left(1+\frac{z_+^2+z_-^2}{2}\right)\sigma_u^2.
\end{aligned} \tag{A.24}$$

Because $\mathrm{Var}(y_1(1))=\mathrm{Var}(x_1(1))+\sigma_u^2$, the expression for $\mathrm{Var}(x_1(1))$ follows immediately.
We can write $x_1(2)$ as

$$x_1(2)=\frac{1}{2\lambda_2}\{(1-\beta_+\Lambda_+-\beta_-\Lambda_-)\widetilde{V}_1-(\beta_-\Lambda_++\beta_+\Lambda_-)\widetilde{V}_2-\Lambda_+u_1(1)-\Lambda_-u_2(1)\}. \tag{A.25}$$

Then

$$\begin{aligned}
\mathrm{Var}(x_1(2)) &= \frac{1}{4\lambda_2^2}\{[(1-\beta_+\Lambda_+-\beta_-\Lambda_-)^2+(\beta_-\Lambda_++\beta_+\Lambda_-)^2\\
&\quad-2\rho(1-\beta_+\Lambda_+-\beta_-\Lambda_-)(\beta_-\Lambda_++\beta_+\Lambda_-)]\sigma_0^2+(\Lambda_+^2+\Lambda_-^2)\sigma_u^2\}.
\end{aligned} \tag{A.26}$$

It can be easily verified that

$$\begin{aligned}
1-\beta_+\Lambda_+-\beta_-\Lambda_- &= \frac{1}{2}\left(\frac{1}{1+z_-^2}+\frac{1}{1+z_+^2}\right),\\
\beta_-\Lambda_++\beta_+\Lambda_- &= \frac{1}{2}\left(\frac{1}{1+z_-^2}-\frac{1}{1+z_+^2}\right),\\
(\Lambda_+^2+\Lambda_-^2)\sigma_u^2 &= \left(\frac{1+\rho}{2}\frac{z_+^2}{1+z_+^2}+\frac{1-\rho}{2}\frac{z_-^2}{1+z_-^2}\right)\sigma_0^2.
\end{aligned} \tag{A.27}$$

Substituting equations (A.27) into (A.26) yields the desired result.

$\text{Cov}(y_1(1),y_2(1))$ can be obtained as follows:

$$\begin{aligned}\text{Cov}(y_1(1),y_2(1)) &= \sigma_0^2(\rho(\beta_+^2+\beta_-^2)+2\beta_+\beta_-)\\ &= \frac{\sigma_0^2\beta_K^2}{4}\left\{\frac{2\rho z_+^2}{1+\rho}+\frac{2\rho z_-^2}{1-\rho}+2\left(\frac{z_+^2}{1+\rho}-\frac{z_-^2}{1-\rho}\right)\right\}\\ &= \left(\frac{z_+^2-z_-^2}{2}\right)\sigma_u^2. \end{aligned}\tag{A.28}$$

In order to prove that the total order flows are not correlated across time, we note that using the symmetry of the solution (specifically that $\text{Var}(y_1(1))=\text{Var}(y_2(1))$), the equation (8) can be written as

$$\Lambda_+ = \frac{\text{Cov}(\widetilde{V}_1,y_1(1))\text{Var}(y_1(1))-\text{Cov}(\widetilde{V}_1,y_2(1))\text{Cov}(y_1(1),y_2(1))}{(\text{Var}(y_1(1)))^2-(\text{Cov}(y_1(1),y_2(1)))^2},\tag{A.29}$$

$$\Lambda_- = \frac{\text{Cov}(\widetilde{V}_1,y_2(1))\text{Var}(y_1(1))-\text{Cov}(\widetilde{V}_1,y_1(1))\text{Cov}(y_1(1),y_2(1))}{(\text{Var}(y_1(1)))^2-(\text{Cov}(y_1(1),y_2(1)))^2}.$$

Substituting $y_1(2)$ into the covariance yields

$$\text{Cov}(y_1(1),y_1(2)) = \frac{1}{2\lambda_2}\{\text{Cov}(\widetilde{V}_1,y_1(1))-\Lambda_+\text{Var}(y_1(1))-\Lambda_-\text{Cov}(y_1(1),y_2(1))\}.\tag{A.30}$$

We just need to show that $\Lambda_+\text{Var}(y_1(1))+\Lambda_-\text{Cov}(y_1(1),y_2(1))=\text{Cov}(\widetilde{V}_1,y_1(1))$, which directly follows from equation (A.29).

To prove the last identity, note that $\Delta P_1(2)$ can be written as

$$\Delta P_1(2)=\lambda_2 y_1(2)+(\Lambda_+-\lambda_1)y_1(1)+\Lambda_- y_2(1).\tag{A.31}$$

Consider now

$$\text{Var}(\Delta P_1(2)) = 2\lambda_2^2\sigma_u^2+(\Lambda_-^2+(\Lambda_+-\lambda_1)^2)\text{Var}(y_1(1))+2\Lambda_-(\Lambda_+-\lambda_1)\text{Cov}(y_1(1),y_2(1)).\tag{A.32}$$

It follows from equation (A.30) that

$$\frac{\lambda_1-\Lambda_+}{\Lambda_-}=\frac{\text{Cov}(y_1(1),y_2(1))}{\text{Var}(y_1(1))}=\frac{z_+^2-z_-^2}{2+z_+^2+z_-^2},\tag{A.33}$$

which upon substitution into equation (A.32) yields the desired result.

$\text{Cov}(y_1(2),y_2(2))$ can be obtained using equations (A.25) and (A.27)

$$\begin{aligned}\text{Cov}(y_1(2),y_2(2)) = \text{Cov}(x_1(2),x_2(2)) &= \frac{\sigma_0^2}{4\lambda_2^2}\Big\{\rho((1-\beta_+\Lambda_+-\beta_-\Lambda_-)^2+(\beta_-\Lambda_+-\beta_+\Lambda_-)^2)\\ &\quad -2(1-\beta_+\Lambda_+-\beta_-\Lambda_-)(\beta_-\Lambda_+-\beta_+\Lambda_-)\\ &\quad +2\beta_K^2\Lambda_+\Lambda_-\Big\}\\ &= \frac{\sigma_0^2}{8\lambda_2^2}\left(\frac{(1+\rho)z_+^2}{(1+z_+^2)^2}-\frac{(1-\rho)z_-^2}{(1+z_-^2)^2}+\frac{1+\rho}{(1+z_+^2)^2}-\frac{1-\rho}{(1+z_-^2)^2}\right),\end{aligned}\tag{A.34}$$

and the result follows immediately.

Finally, we find $\text{Cov}(\Delta P_1(2),\Delta P_2(2))$. Using equation (A.31), we obtain

$$\begin{aligned}\text{Cov}(\Delta P_1(2),\Delta P_2(2)) &= \lambda_2^2\text{Cov}(y_1(2),y_2(2))\\ &\quad +(\Lambda_-^2+(\Lambda_+-\lambda_1)^2)\text{Cov}(y_1(1),y_2(1))\\ &\quad +2\Lambda_-(\Lambda_+-\lambda_1)\text{Var}(y_1(1)).\end{aligned}\tag{A.35}$$

The result follows immediately after applying equation (A.33) in equation (A.35). ||

Theorem A1. *There exists a linear solution to the informed trader's profit maximization problem when A3 is removed characterized by the following parameters.*
At time $t=1$, the informed trader's strategy, $\mathbf{x}(1)=\beta(1)\mathbf{V}$, is characterized by $\beta_{11}(1)=\beta_{22}(1)\equiv\beta_+$ and $\beta_{12}(1)=\beta_{21}(1)\equiv\beta_-$ which are the solutions to the system of non-linear equations:

$$\begin{cases} 2\left(\Lambda_+ - \dfrac{\Lambda_+^2+\Lambda_-^2}{4\lambda_2}\right)\beta_+ + \Lambda_-\left(1-\dfrac{\Lambda_+}{\lambda_2}\right)\beta_- = 1-\dfrac{\Lambda_+}{2\lambda_2}, \\ \Lambda_-\left(1-\dfrac{\Lambda_+}{\lambda_2}\right)\beta_+ + 2\left(\Lambda_+ - \dfrac{\Lambda_+^2+\Lambda_-^2}{4\lambda_2}\right)\beta_- = -\dfrac{\Lambda_-}{2\lambda_2}. \end{cases} \tag{A.36}$$

The market maker pricing function is $\mathbf{P}(1)=\Lambda(\beta(1)\mathbf{V}+\mathbf{u}(1))=\Lambda(1)\mathbf{y}(1)$, with:

$$\Lambda_{11}=\Lambda_{22}\equiv\Lambda_+ = \frac{\psi_+\beta_+ - \psi_-\beta_-}{\psi_+^2-\psi_-^2}, \tag{A.37}$$

$$\Lambda_{12}=\Lambda_{21}\equiv\Lambda_- = \frac{\psi_+\beta_- - \psi_-\beta_+}{\psi_+^2-\psi_-^2},$$

where $\psi_+ = \frac{\beta_K^2}{1-\rho^2}+\beta_+^2+\beta_-^2$ and $\psi_- = -\frac{\rho\beta_K^2}{1-\rho^2}+2\beta_+\beta_-$.
At time $t=2$, the informed trader's strategy is $\mathbf{x}(2)=\beta(2)(\mathbf{V}-\mathbf{P}(1))$ and the market maker pricing function is $\mathbf{P}(2)=\mathbf{P}(1)+\lambda(2)(\mathbf{x}(2)+\mathbf{u}(2))$, with

$$\beta_{11}(2)=\beta_{22}(2)=\frac{1}{\lambda_2}, \tag{A.38}$$

$$\beta_{12}(2)=\beta_{21}(2)=0,$$

$$\lambda_1(2)=\lambda_2(2)\equiv\lambda_2=\sqrt{\frac{1-\beta_+\Lambda_+-\beta_-\Lambda_- - \rho(\beta_+\Lambda_-+\beta_-\Lambda_+)}{4\beta_K^2}}. \tag{A.39}$$

Proof of Theorem A1. The proof follows the logic of Theorem 1 with modifications corresponding to the removal of A3. ‖

Proof of Proposition 1. Because all variances are provided by Lemma 2, we only need to calculate all the necessary covariances in order to prove Proposition 1. We will use the result from Lemma 2 that total order flows are not correlated across time (second relation in equation (A.22)) to prove equation (24)

$$\begin{aligned}\mathrm{Cov}(y_1(1),y_{1,2}(2)) &= \mathrm{Cov}(x_1(1)+u_1(1),x_{1,2}(2))\\ &= \mathrm{Cov}(x_1(1),x_{1,2}(2)) - \frac{1}{2\lambda_2}\mathrm{Cov}(u_1(1),P_{1,2}(1^+))\\ &= \mathrm{Cov}(x_1(1),x_{1,2}(2)) - \frac{1}{2\lambda_2}\mathrm{Cov}(u_1(1),\Lambda_\pm u_1(1)) = 0. \end{aligned} \tag{A.40}$$

It immediately follows that

$$\mathrm{Cov}(x_1(1),x_{1,2}(2)) = \frac{\Lambda_\pm}{2\lambda_2}\sigma_u^2. \tag{A.41}$$

The lead–lag covariances of price changes are given by

$$\begin{aligned}\mathrm{Cov}(\Delta P_1(1),\Delta P_1(2)) &= \lambda_1(\Lambda_+-\lambda_1)\mathrm{Var}(y_1(1))+\lambda_1\Lambda_-\mathrm{Cov}(y_1(1),y_2(1)),\\ \mathrm{Cov}(\Delta P_1(1),\Delta P_2(2)) &= \lambda_1\Lambda_-\mathrm{Var}(y_1(1))+\lambda_1(\Lambda_+-\lambda_1)\mathrm{Cov}(y_1(1),y_2(1)).\end{aligned} \tag{A.42}$$

Combining equations (A.42) and (A.33), we obtain

$$\begin{aligned}\mathrm{Cov}(\Delta P_1(1),\Delta P_1(2)) &= 0,\\ \mathrm{Cov}(\Delta P_1(1),\Delta P_2(2)) &= \frac{\lambda_1\Lambda_-}{\mathrm{Var}(y_1(1))}(\mathrm{Var}(y_1(1))^2-\mathrm{Cov}(y_1(1),y_2(1))^2).\end{aligned} \tag{A.43}$$

Taking into account that

$$\mathrm{Var}(y_1(1))^2-\mathrm{Cov}(y_1(1),y_2(1))^2=(1+z_+^2)(1+z_-^2)\sigma_u^4, \tag{A.44}$$

we immediately obtain

$$\mathrm{Cov}(\Delta P_1(1),\Delta P_2(2))=\frac{2\lambda_1\Lambda_-(1+z_+^2)(1+z_-^2)}{2+z_+^2+z_-^2}\sigma_u^2\geq 0. \tag{A.45}$$

Equation (A.45) combined with

$$\psi_+^2-\psi_-^2=\beta_K^4\frac{(1+z_+^2)(1+z_-^2)}{1-\rho^2}, \tag{A.46}$$

equation (A.33), and equation (A.28) yields equation (23).

The lead–lag covariances of informed order flows and returns are given by

$$\begin{aligned}\mathrm{Cov}(x_1(1),\Delta P_1(2))&=\lambda_2\mathrm{Cov}(x_1(1),x_1(2))+(\Lambda_+-\lambda_1)\mathrm{Var}(x_1(1))+\Lambda_-\mathrm{Cov}(y_1(1),y_2(1))\\&=\frac{\Lambda_+}{2}\sigma_u^2-(\Lambda_+-\lambda_1)\sigma_u^2+\underbrace{(\Lambda_+-\lambda_1)\mathrm{Var}(y_1(1))+\Lambda_-\mathrm{Cov}(y_1(1),y_2(1))}_{0}\\&=\left(\lambda_1-\frac{\Lambda_+}{2}\right)\sigma_u^2.\end{aligned} \tag{A.47}$$

$$\begin{aligned}\mathrm{Cov}(x_1(1),\Delta P_2(2))&=\lambda_2\mathrm{Cov}(x_1(1),x_2(2))+\Lambda_-\mathrm{Var}(x_1(1))+(\Lambda_+-\lambda_1)\mathrm{Cov}(y_1(1),y_2(1))\\&=\frac{\Lambda_-}{2}\sigma_u^2-\Lambda_-\sigma_u^2+\Lambda_-\mathrm{Var}(y_1(1))+(\Lambda_+-\lambda_1)\mathrm{Cov}(y_1(1),y_2(1))\\&=-\frac{\Lambda_-}{2}\sigma_u^2+\frac{1}{\lambda_1}\mathrm{Cov}(\Delta P_1(1),\Delta P_2(2)),\end{aligned} \tag{A.48}$$

which in combination with equation (A.33) yields

$$\mathrm{Cov}(x_1(1),\Delta P_1(2))=\Lambda_+\sigma_u^2\left(\frac{1}{2}+\frac{\Lambda_-(z_+^2-z_-^2)}{\Lambda_+(2+z_+^2+z_-^2)}\right)>0, \tag{A.49}$$

$$\mathrm{Cov}(x_1(1),\Delta P_2(2))=\Lambda_-\sigma_u^2\left(\frac{1}{2}+\frac{(z_++z_-)^2}{2+z_+^2+z_-^2}\right)\geq 0. \tag{A.50}$$

Combining the covariances with the results of Lemma 2 completes the proof. ||

Proof of Remark 1. We now demonstrate that the A3 at time 1 is behind the positive portfolio autocovariance in Proposition 2. If A3 is removed at time 1, then the first-period price vector is equal to

$$\mathbf{P}(1)=\mathrm{E}[\widetilde{\mathbf{V}}|y_1(1),y_2(1)]=\begin{pmatrix}\Lambda_{11}&\Lambda_{12}\\\Lambda_{12}&\Lambda_{22}\end{pmatrix}\begin{pmatrix}y_1(1)\\y_2(1)\end{pmatrix}=\Lambda\mathbf{y}(1), \tag{A.51}$$

where Λ_{ij} are given by equation (8). The second period price vector is

$$\mathbf{P}(2)=\mathbf{P}(1)+\lambda(2)\mathbf{y}(2), \tag{A.52}$$

where matrix $\lambda(2)$ is diagonal as in equation (14). We also have that $y(2)=\beta(2)(\mathbf{V}-\mathbf{P}(1))+u(2)$, where $\beta(2)$ is also diagonal. It is straightforward to establish that

$$\begin{aligned}\mathrm{Cov}(\Delta P_i(1),\Delta P_i(2))&=\lambda_i(2)(\Lambda_{i1}\mathrm{Cov}(y_1(1),y_i(2))+\Lambda_{i2}\mathrm{Cov}(y_2(1),y_i(2))),\\\mathrm{Cov}(\Delta P_i(1),\Delta P_{i\neq j}(2))&=\lambda_{i\neq j}(2)(\Lambda_{i1}\mathrm{Cov}(y_1(1),y_{i\neq j}(2))+\Lambda_{i2}\mathrm{Cov}(y_2(1),y_{i\neq j}(2))).\end{aligned} \tag{A.53}$$

All we need to show is that autocorrelations in total order flows are zero. Consider

$$\begin{aligned}\mathrm{Cov}(y_1(1),y_i(2))&=\beta_i(2)(\mathrm{Cov}(\widetilde{V}_i,y_1(1))-\Lambda_{i1}\mathrm{Var}(y_1(1))-\Lambda_{i2}\mathrm{Cov}(y_1(1),y_2(1))),\\\mathrm{Cov}(y_2(1),y_i(2))&=\beta_i(2)(\mathrm{Cov}(\widetilde{V}_i,y_2(1))-\Lambda_{i1}\mathrm{Cov}(y_1(1),y_2(1))-\Lambda_{i2}\mathrm{Var}(y_1(1))),\end{aligned} \tag{A.54}$$

for $i=1,2$. It directly follows from equation (8) that

$$\mathrm{Cov}(\widetilde{V}_i, y_1(1)) = \Lambda_{i1}\mathrm{Var}(y_1(1)) + \Lambda_{i2}\mathrm{Cov}(y_1(1), y_2(1)),$$
$$\mathrm{Cov}(\widetilde{V}_i, y_2(1)) = \Lambda_{i2}\mathrm{Var}(y_2(1)) + \Lambda_{i1}\mathrm{Cov}(y_1(1), y_2(1)), \tag{A.55}$$

and we immediately obtain that the total order flow autocovariances and, as a consequence, securities' returns autocovariances, are all zero. ‖

Proof of Proposition 2. The results for $\mathrm{Cov}(\Delta P_p(1), \Delta P_p(2))$ and $\mathrm{Cov}(\Delta P_p(1), x_p(2))$ follow immediately from Proposition 1. Next we consider

$$\mathrm{Cov}(x_p(1), x_p(2)) = \frac{1}{2}(\mathrm{Cov}(x_1(1), x_1(2)) + \mathrm{Cov}(x_1(1), x_2(2))), \tag{A.56}$$

and the result follows when we use equations(A.41) in equation (A.56). Finally, consider

$$\mathrm{Cov}(x_p(1), \Delta P_p(2)) = \frac{1}{2}[\mathrm{Cov}(x_1(1), \Delta P_1(2)) + \mathrm{Cov}(x_1(1), \Delta P_2(2))], \tag{A.57}$$

and the result follows immediately. ‖

The following Lemma is useful to prove Proposition 3.

Lemma A3. *Portfolio informed order flows have the following variances:*

$$\mathrm{Var}(x_p(1)) = \frac{z_+^2}{2}\sigma_u^2, \tag{A.58}$$
$$\mathrm{Var}(x_p(2)) = \frac{1}{2}(\sigma_u^2 + \mathrm{Cov}(x_1(2), x_2(2))).$$

Portfolio price changes have the following variances:

$$\mathrm{Var}(\Delta P_p(1)) = \lambda_1^2\left(\frac{\sigma_u^2}{2} + \mathrm{Var}(x_p(1))\right), \tag{A.59}$$
$$\mathrm{Var}(\Delta P_p(2)) = \lambda_2^2\left(\sigma_u^2 + \frac{1}{2}\mathrm{Cov}(y_1(2), y_2(2))\right) + \frac{2\Lambda_-^2(1+z_+^2)(1+z_-^2)^2}{(2+z_+^2+z_-^2)^2}.$$

Portfolio price changes and informed order flows have the following covariance:

$$\mathrm{Cov}(x_p(1), \Delta P_p(1)) = \lambda_1 \mathrm{Var}(x_p(1)). \tag{A.60}$$

Proof of Lemma A3. The variance of $x_p(1)$ can be calculated as follows:

$$\mathrm{Var}(x_p(1)) = \frac{1}{2}(\mathrm{Var}(x_1(1)) + \mathrm{Cov}(x_1(1), x_2(1)))$$
$$= \frac{\sigma_u^2}{2}\left(\frac{z_+^2+z_-^2}{2} + \frac{z_+^2-z_-^2}{2}\right). \tag{A.61}$$

$\mathrm{Var}(x_p(2))$ is equal to

$$\mathrm{Var}(x_p(2)) = \frac{1}{2}(\underbrace{\mathrm{Var}(x_1(2))}_{\sigma_u^2} + \mathrm{Cov}(x_1(2), x_2(2))), \tag{A.62}$$

and the result follows. The other proofs are straightforward except for $\mathrm{Var}(\Delta P_p(2))$, which we prove next.

$$\mathrm{Var}(\Delta P_p(2)) = \frac{1}{2}(\mathrm{Var}(\Delta P_1(2)) + \mathrm{Cov}(\Delta P_1(2), \Delta P_2(2))) = \lambda_2^2\left(\sigma_u^2 + \frac{1}{2}\mathrm{Cov}(y_1(2), y_2(2))\right)$$
$$+ \Lambda_-^2\left(\frac{\lambda_1 - \Lambda_+}{\Lambda_-} - 1\right)^2(\mathrm{Var}(y_1(1)) + \mathrm{Cov}(y_1(1), y_2(1))). \tag{A.63}$$

The result follows after substituting equation (A.33) into equation (A.32). ‖

Proof of Proposition 3. The VAR coefficients b_{11} and b_{12} are found from

$$\begin{pmatrix} b_{11} \\ b_{12} \end{pmatrix} = \underbrace{\begin{pmatrix} \mathrm{Var}(\Delta P_p(1)) & \mathrm{Cov}(x_p(1), \Delta P_p(1)) \\ \mathrm{Cov}(x_p(1), \Delta P_p(1)) & \mathrm{Var}(x_p(1)) \end{pmatrix}^{-1}}_{A} \begin{pmatrix} \mathrm{Cov}(\Delta P_p(1), \Delta P_p(2)) \\ \mathrm{Cov}(x_p(1), \Delta P_p(2)) \end{pmatrix}. \tag{A.64}$$

We need to calculate

$$\det(A)=\frac{1}{2}\lambda_1^2\sigma_u^2\mathrm{Var}(x_p(1)). \tag{A.65}$$

Next we find b_{11}

$$\begin{aligned} b_{11}&=\frac{1}{\det(A)}(\mathrm{Cov}(\Delta P_p(1),\Delta P_p(2))\mathrm{Var}(x_p(1))-\mathrm{Cov}(x_p(1),\Delta P_p(2))\mathrm{Cov}(x_p(1),\Delta P_p(1)))\\ &=\frac{\mathrm{Cov}(\Delta P_p(1),\Delta P_p(2))-\lambda_1\mathrm{Cov}(x_p(1),\Delta P_p(2))}{\frac{1}{2}\lambda_1^2\sigma_u^2}=\frac{\Lambda_++\Lambda_-}{2\lambda_1}-1. \end{aligned} \tag{A.66}$$

Using relation (A.33), which can be rewritten as

$$\lambda_1-\frac{\Lambda_+}{2}=\frac{\Lambda_+}{2}+\frac{\Lambda_-(z_+^2-z_-^2)}{2+z_+^2+z_-^2},$$

in equation (A.66), we obtain

$$b_{11}=-\frac{1}{2\lambda_1}\left(\frac{\Lambda_+-\Lambda_-}{2}+\frac{\Lambda_-(z_+^2-z_-^2)}{2+z_+^2+z_-^2}\right)<0.$$

b_{12} can be found analogously

$$\begin{aligned} b_{12}&=\frac{1}{\det(A)}(\mathrm{Cov}(x_p(1),\Delta P_p(2))\mathrm{Var}(\Delta P_p(1))-\mathrm{Cov}(\Delta P_p(1),\Delta P_p(2))\mathrm{Cov}(x_p(1),\Delta P_p(1)))\\ &=\frac{1+z_+^2}{z_+^2}\left(\frac{\Lambda_++\Lambda_-}{2}+\frac{2\Lambda_-(z_+^2+z_+z_-)}{2+z_+^2+z_-^2}\right)-\frac{2\Lambda_-(1+z_+^2)(1+z_-^2)}{2+z_+^2+z_-^2}\\ &=\left(1+\frac{1}{z_+^2}\right)\left(\frac{\Lambda_++\Lambda_-}{2}+\frac{2\Lambda_-z_+z_-(1-z_+z_-)}{2+z_+^2+z_-^2}\right)>0. \end{aligned} \tag{A.67}$$

Alternatively, it can be rewritten as

$$\begin{aligned} b_{12}&=\frac{2}{\sigma_u^2}\left[\left(1+\frac{1}{z_+^2}\right)\mathrm{Cov}(x_p(1),\Delta P_p(2))-\frac{1}{\lambda_1}\mathrm{Cov}(\Delta P_p(1),\Delta P_p(2))\right]\\ &=\left(1+\frac{1}{z_+^2}\right)\left(\lambda_1-\frac{\Lambda_++\Lambda_-}{2}\right)+\frac{2}{\lambda_1\sigma_u^2z_+^2}\mathrm{Cov}(\Delta P_1(1),\Delta P_2(2)). \end{aligned}$$

We can now find b_{22} and b_{21} that are equal to

$$\begin{pmatrix} b_{21}\\ b_{22}\end{pmatrix}=\begin{pmatrix}\mathrm{Var}(\Delta P_p(1)) & \mathrm{Cov}(x_p(1),\Delta P_p(1))\\ \mathrm{Cov}(x_p(1),\Delta P_p(1)) & \mathrm{Var}(x_p(1))\end{pmatrix}\begin{pmatrix}\mathrm{Cov}(\Delta P_p(1),x_p(2))\\ \mathrm{Cov}(x_p(1),x_p(2))\end{pmatrix}. \tag{A.68}$$

Therefore,

$$\begin{aligned} b_{21}&=\frac{1}{\det(A)}(\underbrace{\mathrm{Cov}(\Delta P_p(1),x_p(2))\mathrm{Var}(x_p(1))}_{0}-\mathrm{Cov}(x_p(1),x_p(2))\mathrm{Cov}(x_p(1),\Delta P_p(1)))\\ &=-\frac{\Lambda_++\Lambda_-}{2\lambda_1\lambda_2}. \end{aligned} \tag{A.69}$$

Finally,

$$\begin{aligned} b_{22}&=\frac{1}{\det(A)}(\mathrm{Cov}(x_p(1),x_p(2))\mathrm{Var}(\Delta P_p(1))-\underbrace{\mathrm{Cov}(\Delta P_p(1),x_p(2))\mathrm{Cov}(x_p(1),\Delta P_p(1)))}_{0}\\ &=\left(1+\frac{1}{z_+^2}\right)\frac{\Lambda_++\Lambda_-}{2\lambda_2}. \end{aligned} \tag{A.70}$$

The result follows after using the definition of z_+ (equation (A.12)) in equation (A.70). ‖

Derivation of equation (34). Using Lemmas 2 and 3, it is straightforward to verify that

$$\begin{aligned}\mathrm{Var}(\widehat{x}_p(1)) &= (1-M)^2\mathrm{Var}(x_p(1)) + M^2\frac{\sigma_u^2}{2},\\ \mathrm{Cov}(\widehat{x}_p(1),\Delta P_p(1)) &= (1-M)\mathrm{Cov}(x_p(1),\Delta P_p(1)) - M\lambda_1\frac{\sigma_u^2}{2},\\ \mathrm{Cov}(\widehat{x}_p(1),\Delta P_p(2)) &= (1-M)\mathrm{Cov}(x_p(1),\Delta P_p(2)) + M\left(\lambda_1 - \frac{\Lambda_+ + \Lambda_-}{2}\right)\frac{\sigma_u^2}{2}.\end{aligned} \tag{A.71}$$

Next we establish that

$$\begin{aligned}\det(\widehat{A}) &= \mathrm{Var}(\Delta P_p(1))\mathrm{Var}(\widehat{x}_p(1)) - \mathrm{Cov}(\widehat{x}_p(1),\Delta P_p(1))^2\\ &= (1-M)^2\det(A) + \frac{M^2\lambda_1^2}{4}(\sigma_u^2 + 2\mathrm{Var}(x_p(1))) + M(1-M)\lambda_1^2\mathrm{Var}(x_p(1)) - \frac{M^2\lambda_1^2\sigma_u^2}{4}\\ &= (1-M)^2\det(A) + M\left(1-\frac{M}{2}\right)\underbrace{\lambda_1^2\mathrm{Var}(x_p(1))}_{2\det(A)} = \det(A).\end{aligned} \tag{A.72}$$

Finally, we find $\widehat{b}_{11}$

$$\begin{aligned}\widehat{b}_{11} &= \frac{1}{\det(A)}(\mathrm{Cov}(\Delta P_p(1),\Delta P_p(2))\mathrm{Var}(\widehat{x}_p(1)) - \mathrm{Cov}(\widehat{x}_p(1),\Delta P_p(2))\mathrm{Cov}(\widehat{x}_p(1),\Delta P_p(1)))\\ &= \frac{(1-M)^2 b_{11}\det(A) + \frac{M\lambda_1\sigma_u^2}{2}\mathrm{Cov}(x_p(1),\Delta P_p(2)) + M(1-M)b_{11}\det(A)}{\det(A)},\end{aligned} \tag{A.73}$$

and the result follows.

In order to calculate $\widehat{b}_{21}$, we need to find $\mathrm{Cov}(\widehat{x}_p(1),\widehat{x}_p(2))$

$$\mathrm{Cov}(\widehat{x}_p(1),\widehat{x}_p(2)) = (1-M)\mathrm{Cov}(x_p(1),x_p(2)), \tag{A.74}$$

yielding the expression for $\widehat{b}_{21}$

$$\begin{aligned}\widehat{b}_{21} &= \frac{1}{\det(A)}(\underbrace{\mathrm{Cov}(\Delta P_p(1),\widehat{x}_p(2))\mathrm{Var}(\widehat{x}_p(1))}_{0} - \mathrm{Cov}(\widehat{x}_p(1),\widehat{x}_p(2))\mathrm{Cov}(\widehat{x}_p(1),\Delta P_p(1)))\\ &= \frac{(1-M)^2 b_{21}\det(A) + \frac{M(1-M)\lambda_1\sigma_u^2}{2}\mathrm{Cov}(x_p(1),x_p(2))}{\det(A)}.\end{aligned} \tag{A.75}$$

Acknowledgments. We thank Bruno Biais, Larry Glosten, Tarun Ramadorai, Mark Rubinstein, Jules van Binsbergen, two anonymous referees, and seminar participants at the Commodity Futures Trading Commission, Dauphine University, Erasmus University, Hebrew University, HEC Lausanne, Simon Fraser University, Stanford University, Tel-Aviv University, Tilburg University, University of California, Berkeley, University of California, San Diego, and University of Tokyo for helpful comments.

REFERENCES

ADMATI, A. (1985), "A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Markets", *Econometrica*, **53**, 629–657.

ANDERSON, R. M., EOM, K. S., HAHN, S. B. and PARK, J. H. (2008), "Stock Return Autocorrelation Is Not Spurious" (Working Paper, University of California, Berkeley).

ASQUITH, A., OMAN, R. and SAFAYA, C. (2010), "Short Sales and Trade Classification Algorithms", *Journal of Financial Markets*, **13**, 157–173.

BADRINATH, S. G., KALE, J. R. and NOE, T. H. (1995), "Of Shepherds, Sheep, and the Crossautocorrelations in Equity Returns", *Review of Financial Studies*, **8**, 401–430.

BERNHARDT, D. and DAVIES, R. (2008), "The Impact of Nonsynchronous Trading on Differences in Portfolio Cross-Autocorrelations" (Working Paper, University of Illinois).

BERNHARDT, D. and MAHANI, R. (2007), "Asymmetric Information and Stock Return Cross-Autocorrelations", *Economics Letters*, **62**, 14–22.

BERNHARDT, D. and TAUB, D. (2006), "Kyle v. Kyle ('85 v. '89)", *Annals of Finance*, **2**, 23–38.

BERNHARDT, D. and TAUB, D. (2008), "Cross-Asset Speculation in Stock Markets", *Journal of Finance*, **63**, 2385–2427.

BIAIS, B., GLOSTEN, L. and SPATT, C. (2005), "Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications", *Journal of Financial Markets*, **8**, 217–264.

BIAIS, B., HOMBERT, J. and WEILL, P. (2010), "Trading and Liquidity with Limited Cognition" (IDEI Working Paper No. 665, University of Toulouse).

BIAIS, B., MARTIMORT, D. and ROCHET, J. C. (2000), "Continuous Auctions and Informed Trading", *Econometrica*, **53**, 1315–1335.

BOEHMER, E. and KELLEY, E. (2009), "Institutional Investors and the Informational Efficiency of Prices", *Review of Financial Studies*, **22**, 3563–3594.

BOEHMER, E. and WU, J. (2008), "Order Flow and Prices" (Working Paper, University of Georgia).

BOUDOUKH, J., RICHARDSON, M. and WHITELAW, R. (1994), "A Tale of Three Schools: Insights on Autocorrelation of Short-Horizon Returns", *Review of Financial Studies*, **7**, 539–573.

BRENNAN, M. J., JAGADEESH, N. and SWAMINATHAN, B. (1993), "Investment Analysis and the Adjustment of Stock Prices to Common Information", *Review of Financial Studies*, **6**, 799–824.

CABALLE, J. and KRISHNAN, M. (1994), "Imperfect Competition in a Multi-Security Market with Risk Neutrality", *Econometrica*, **62**, 695–704.

CAMPBELL, J. Y., LO, A. W. and MACKINLAY, A. C. (1997), *The Econometrics of Financial Markets* (New Jersey: Princeton University Press).

CESPA, G. and FOUCAULT, T. (2010), "Dealer Attention, Liquidity Spillovers and Endogenous Market Segmentation" (Working Paper, HEC Paris).

CHAN, K. (1993), "Imperfect Information and Cross-Autocorrelation Among Stock Prices", *Journal of Finance*, **48**, 1211–1230.

CHORDIA, T., ROLL, R. and SUBRAHMANYAM, A. (2005), "Evidence on the Speed of Convergence to Market Efficiency", *Journal of Financial Economics*, **76**, 271–292.

CHORDIA, T., SARKAR, A. and SUBRAHMANYAM, A. (2011), "Liquidity Dynamics and Cross-Autocorrelations", *Journal of Financial and Quantitative Analysis*, **46**, 709–736.

CHORDIA, T. and SWAMINATHAN, B. (2000), "Trading Volume and Cross-Autocorrelations in Stock Returns", *Journal of Finance*, **55**, 913–936.

CONRAD, J. and KAUL, G. (1988), "Time-Variation in Expected Returns", *Journal of Business*, **61**, 409–425.

DENNERT, J. (1993), "Price Competition Between Market Makers", *Review of Economic Studies*, **60**, 735–751.

HASBROUCK, J. (1991), "Measuring the Information Content of Stock Trades", *Journal of Finance*, **46**, 179–207.

HENDERSHOTT, T., JONES, C. and MENKVELD, A. (2011), "Does Algorithmic Trading Improve Liquidity?" *Journal of Finance*, **66**, 1–33.

HOU, K. (2007), "Industry Information Diffusion and the Lead-Lag Effect in Stock Returns", *Review of Financial Studies*, **20**, 1113–1138.

KING, B. (1966), "Market and Industry Factors in Stock Price Behavior", *Journal of Business*, **39**, 139–190.

KYLE, A. S. (1985), "Continuous Auctions and Informed Trading", *Econometrica*, **53**, 1315–1335.

KYLE, A. S. (1989), "Informed Speculation with Imperfect Competition", *Review of Economic Studies*, **56**, 317–355.

LO, A. and MACKINLAY, A. C. (1988), "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test", *Review of Financial Studies*, **1**, 41–66.

LO, A. and MACKINLAY, A. C. (1990a), "When Are Contrarian Profits due to Stock Market Overreaction?" *Review of Financial Studies*, **3**, 175–205.

LO, A. and MACKINLAY, A. C. (1990b), "An Econometric Analysis of Nonsynchronous Trading", *Journal of Econometrics*, **45**, 181–211.

PASQUARIELLO, P. and VEGA, C. (2011), "Strategic Cross-Trading in the U.S. Stock Market" (Working Paper, University of Michigan).

ROCHET, J. C. and VILA, J. L. (1994), "Insider Trading without Normality", *Review of Economic Studies*, **61**, 131–152.

SIAS, R. W. and STARKS, L. T. (1997), "Return Autocorrelation and Institutional Investors", *Journal of Financial Economics*, **46**, 103–131.