Review of Finance (2015) 19: pp. 229–282
doi:10.1093/rof/rft055
Advance Access publication: December 24, 2013

Strategic Cross-Trading in the U.S. Stock Market*

PAOLO PASQUARIELLO[1] and CLARA VEGA[2]
[1]*Ross School of Business and* [2]*the Federal Reserve Board of Governors*

Abstract. We model and test for the role of heterogeneously informed, strategic multi-asset speculation for cross-price impact—the impact of trades in one asset on the prices of other (even unrelated) assets—in the U.S. stock market. Our investigation of the trading activity in New York Stock Exchange (NYSE) and National Association of Securities Dealers Automated Quotation System (NASDAQ) stocks between 1993 and 2004 reveals that, consistent with our model, (1) daily order imbalance in one industry or random stock has a significant, persistent, and robust impact on daily returns of other (even unrelated) industries or random stocks; (2) cross-price impact is often negative; and (3) both direct (i.e., an asset's own) and absolute (i.e., unsigned) cross-price impact are smaller when speculators are more numerous, greater when market-wide dispersion of beliefs is higher, and greater among stocks dealt by the same specialist.

JEL Classification: D82, G14

1. Introduction

What moves stock prices? A large body of research relates this fundamental question in financial economics to market frictions (e.g., liquidity, transaction costs, financing and short-selling constraints, information asymmetry and heterogeneity).[1] Recent studies also find strong evidence of cross-stock

* The authors thank the Mitsui Financial Research Center for financial support, Hien Tran for outstanding research assistance, and Thierry Foucault (the editor), an anonymous referee, Robert Battalio, Alessandro Beber, Paul Bennett, Sreedhar Bharath, Tarun Chordia, Shane Corwin, Diego Garcia, Robin Greenwood, Robert Jennings, Kenneth Lee, Bruce Lehmann, Bruce Mizrach, Pamela Moulton, Amiyatosh Purnanandam, Uday Rajan, Gideon Saar, Avanidhar Subrahmanyam, Hayong Yun, and seminar participants at the University of North Carolina, University of Amsterdam, Vanderbilt University, University of Toronto, University of Notre Dame, EPFL, 2008 NBER Market Microstructure meetings, Second Erasmus Liquidity Conference, and 2010 AFA meetings for comments.

[1] See Amihud and Mendelson (1986), Constantinides (1986), De Long *et al.* (1990), Brennan and Subrahmanyam (1996), Brennan, Chordia, and Subrahmanyam (1998), Vayanos (1998), Shleifer (2000), Amihud (2002), Huang (2003), Pastor and Stambaugh

linkages (Hasbrouck and Seppi, 2001; Hartford and Kaul, 2005; Andrade, Chang, and Seasholes, 2008; Tookes, 2008). However, the effect of those frictions on the process of price co-formation in equity markets remains poorly understood. Motivated by this literature, we undertake a novel investigation of the role of heterogeneously informed, strategic speculation for the *cross-price impact* of order flow (the impact of trading activity in one asset on the prices of other, even *unrelated* assets) in the U.S. stock market.

Our empirical analysis is guided by a multi-asset sequential trade model of speculation based on Kyle (1985) and Caballé and Krishnan (1994). This framework is analytically tractable and allows us to explicitly illustrate the relationship between cross-price impact and strategic, informed multi-asset trading activity. The basic intuition of our model is as follows. In an interconnected economy—where most assets are fundamentally related—uninformed market-makers (MMs) attempt to learn about the liquidation value of one asset from order flow in other assets. Thus, imperfectly competitive speculators, when better informed about an asset, optimally trade strategically in many assets (even unrelated ones). They do so for two reasons: First, to attenuate the dissipation of their information advantage in that asset—that is, the *direct price impact* of trading in that asset; second, to mitigate the trading costs of their strategy. Being rational, the MMs take into account speculators' strategies when setting prices. In equilibrium, the speculators' cross-trading and subsequent MMs' cross-inference from it leads to cross-price impact (even among unrelated assets). In this setting, we are to our knowledge the first to show that, because of the ensuing adverse selection risk for the MMs: (1) equilibrium informational cross-price impact may be *negative* even when asset payoff covariances are not; and (2) both direct and absolute (i.e., unsigned) cross-price impact are decreasing in the number of speculators and increasing in the heterogeneity of their private information.

We test our model's implications in the U.S. stock market by analyzing the Trade and Quote (TAQ) database—the most comprehensive sample of trading activity in the NYSE and the NASDAQ—between 1993 and 2004. For the sake of parsimony (given the large number of stocks in each year of the sample period), we estimate the cross-price impact of order flow either among 10 industry-sorted stock portfolios or among a large number of randomly selected stock pairs. Overall, we find the ensuing empirical evidence to support the main predictions of our model.

(2003), Acharya and Pedersen (2005), Duffie, Garleanu, and Pedersen (2005), and Sadka and Scherbina (2007).

First, we show that measures of persistent (i.e., informational) cross-industry and cross-stock price impact are often negative and both economically and statistically significant—averaging more than a third of the corresponding measures of direct price impact—even among less closely related industries and stocks. For instance, we estimate that a 1 standard deviation shock to an industry's net order flow affects *another* industry's daily stock returns by an average of 10 basis points (versus an average of 28 basis points from a similar shock to that industry's own order flow). We further find that a 1 standard deviation shock to order flow in one randomly selected stock statistically significantly affects the returns of *another* randomly selected stock often, by an average of 15 basis points (versus an average direct impact of 43 basis points), and even when their absolute earnings correlations are low.

Our evidence of cross-asset informational effects is robust to controlling for marketwide trading activity and price fluctuations (King and Wadhwani, 1990; Hasbrouck and Seppi, 2001), inventory management considerations (Chordia, Roll, and Subrahmanyam, 2000), and any public direct and cross-asset information already embedded in past prices (Chan, 1993; Chordia, Sarkar, and Subrahmanyam, 2011; Boulatov, Hendershott, and Livdan, 2013). Our evidence is also robust to several alternative empirical specifications and to explicitly controlling for alternative channels of trade and price co-formation in the literature (correlated information, portfolio rebalancing, correlated liquidity, and price observability), absent from our model by construction. For instance, we report that estimates of persistent direct and cross-price impact among random NYSE stock pairs dealt by the same specialist—where cross-order flow observability, cross-inference, and strategic cross-trading are likely to be most intense, as in our model—are on average 16 and 2 basis points higher, respectively, than among random NYSE stock pairs dealt by different specialists, ceteris paribus for their pairwise earnings correlations. Finally, our evidence is stable over the sample period and qualitatively similar across alternative trading platforms (i.e., among both NYSE-only and NASDAQ-only stocks).

Further support for our model comes from testing its additional, unique predictions. In particular, we document that, consistent with our model, direct and absolute cross-price impact are higher when speculators are less numerous in the market or when various measures of the dispersion of their beliefs are higher. For example, we find that daily Telecom stock returns increase by an average of 15 basis points from a 1 standard deviation shock to order flow in Nondurables stocks when the number of speculators is low or market-wide dispersion of beliefs is high, but are insensitive to trading activity in Nondurable stocks otherwise. This is despite the fact that correlation between the quarterly earnings of these industries is low (0.152) and statistically insignificant over the sample period. We also find that when market-wide information heterogeneity is high (speculators are less numerous), a 1 standard deviation shock to a random stock's order flow affects its daily returns and the daily returns of another randomly selected stock (if statistically significantly) by an average of 25 (9) and 16 (9) basis points more, respectively, than when information heterogeneity is low (speculators are more numerous).

Our findings have important implications for the study of the process of price formation in financial markets. First, our analysis indicates that (1) the price dynamics of one traded asset may be both economically and statistically significantly related to trading activity in other assets; and (2) this relationship may at least partly stem from the endogenous strategic cross-trading activity of sophisticated speculators and the informational environment in which that activity takes place. These insights are likely to be relevant for most financial markets (e.g., for foreign stocks, currencies, and bonds), and even in light of recent innovations to trading mechanisms, institutions, and regulations both enhancing order flow observability and making strategic cross-trading more viable. Second, our analysis proposes that the liquidity of a traded asset cannot be assessed in isolation, for example, by focusing exclusively on that asset's own trading activity (as common in the literature). Lastly, these results may be crucial to understanding the extent and dynamics of asset price co-movement, an issue at the center of a burgeoning literature (e.g., see the surveys in Veldkamp, 2006; Pasquariello, 2007) and relevant to both risk management and portfolio management.

The article is organized as follows. In Section 2, we construct our model. In Section 3, we describe the data. In Section 4, we present the empirical results. In Section 5, we perform robustness checks of our inference. In Section 6, we review the related literature on cross-stock linkages. We conclude in Section 7.

2. Theoretical Model

In this section, we develop a theoretical model to guide our empirical investigation of the informational role of direct and cross-asset strategic trading in the U.S. stock market. We first describe a parsimonious model of multi-asset sequential trading based on Kyle (1985) and Caballé and Krishnan (1994). Then, we derive closed-form solutions (and testable implications) for the equilibrium prices, market liquidity, and trading strategies as a function of the number of sophisticated market participants and the dispersion of their beliefs.[2] All proofs are in the Appendix.

2.1 THE BASIC SETTING

The model consists of a two-period economy in which N risky assets are exchanged. Trading occurs only at the end of the first period ($t=1$). At the end of the second period ($t=2$), the payoffs of the risky assets, an $N \times 1$ multivariate normally distributed (MND) random vector v with mean P_0 and nonsingular covariance matrix Σ_v, are realized. The economy is populated by three types of risk neutral traders: a discrete number M of informed traders (labeled speculators), liquidity traders, and perfectly competitive MMs. All traders know the structure of the economy and the decision process leading to order flow and prices.

At $t=0$, there is neither information asymmetry about v nor trading. Sometime between $t=0$ and $t=1$, each speculator m receives a private and noisy signal of v, S_{vm}. We assume that each vector S_{vm} is drawn from a MND with mean P_0 and covariance matrix Σ_s and that, for any two speculators m and k, $cov(v, S_{vm}) = cov(v, S_{vk}) = cov(S_{vm}, S_{vk}) = \Sigma_v$. We further parametrize the degree of diversity among speculators' private information by imposing that $\Sigma_s = \frac{1}{\rho}\Sigma_v$ and $\rho \in (0, 1)$, as in Pasquariello and Vega (2009). These assumptions imply that speculator m's information advantage about v at $t=1$, before trading with the MMs, is given by

$$\delta_m \equiv E(v|S_{vm}) - P_0 = \rho(S_{vm} - P_0), \tag{1}$$

where $var(\delta_m) \equiv \Sigma_\delta = \rho\Sigma_v$ is nonsingular. It then follows that any two vectors δ_m and δ_k have a joint MND and $cov(\delta_m, \delta_k) \equiv \Sigma_c = \rho\Sigma_\delta$, a symmetric positive definite (SPD) matrix. Therefore, $E(\delta_k|S_{vm}) = \rho\delta_m$ and ρ can be interpreted as the correlation between any two information endowments δ_m and δ_k: The lower (higher) is ρ, the more (less) *heterogeneous*—that is, the less (more) correlated and, of course, precise—is speculators' private information about v.

[2] Holden and Subrahmanyam (1992), Foster and Viswanathan (1996), Back, Cao, and Willard (2000), and Pasquariello and Vega (2007) develop one-asset extensions of Kyle (1985) to link market liquidity to the trading activity of heterogeneously informed traders. Pasquariello (2007) characterizes the circumstances under which such activity may magnify equilibrium price comovement in Caballé and Krishnan's (1994) multi-trader, multi-asset generalization of Kyle (1985).

At $t = 1$, both speculators and liquidity traders move first and submit their orders to the MMs before the price vector P_1 has been set. We define the vector of market orders of speculator m to be X_m. Thus, her profit is given by $\pi_m(X_m, P_1) = X'_m(v - P_1)$. Liquidity traders generate a vector of random demands z, MND with mean $\underline{0}$ (a zero vector) and nonsingular covariance matrix Σ_z. For simplicity, we impose that noise trading z has identical variance and is independent across assets ($\Sigma_z = \sigma_z^2 I$) and from any other random vector. MMs do not receive any information, but observe the net order flow for each asset $\omega_1 = \sum_{m=1}^{M} X_m + z$ and set the market-clearing prices $P_1 = P_1(\omega_1)$.

2.1.a. *Equilibrium*

Consistent with Caballé and Krishnan (1994), we define a Bayesian Nash equilibrium of this economy as a set of $M+1$ vector functions $X_1(\cdot), \ldots, X_M(\cdot)$, and $P_1(\cdot)$ such that the following two conditions hold:

(1) *Profit maximization:* $X_m(S_{vm}) = \arg\max E(\pi_m | S_{vm})$;

(2) *Semi-strong market efficiency:* $P_1(\omega_1) = E(v|\omega_1)$.

Proposition 1 characterizes the unique linear equilibrium for this economy.

Proposition 1

There exists a unique linear equilibrium given by the price function

$$P_1 = P_0 + \Lambda\omega_1 \tag{2}$$

and by each speculator m's demand strategy

$$X_m = \frac{1}{2 + (M-1)\rho}\Lambda^{-1}\delta_m, \tag{3}$$

where

$$\Lambda = \frac{\sqrt{M\rho}}{[2 + (M-1)\rho]\sigma_z}\Sigma_v^{1/2} \tag{4}$$

is an SPD matrix.

The optimal trading strategy of each speculator depends on the private information she receives about v (δ_m) as well as on the depth of the market (Λ^{-1}). Speculators are imperfectly competitive. Hence, albeit risk neutral, they exploit their information advantage in each market cautiously ($|X_m(n)| < \infty$) to avoid dissipating their informational advantage with their trades, as in the single-asset setting of Kyle (1985). For the same purpose, speculators also trade strategically across assets $\left(\frac{\partial X_m(j)}{\partial \delta_m(n)} \neq 0\right)$.

Intuitively, the MMs know the structure of the economy (the covariance matrix Σ_v). Hence, unless all securities' terminal payoffs are *fundamentally unrelated* (i.e., unless Σ_v is diagonal), they rationally use the order flow for each asset to learn about the liquidation values of other assets when setting the market-clearing price vector $P_1 \left(\frac{\partial P_1(n)}{\partial \omega_1(j)} \neq 0\right)$. Speculators are aware of this learning process, labeled cross-inference. Thus, they strategically place their trades in many assets—rather than independently trading in each asset—to limit the amount of information divulged by their market orders. As a result of this effort, labeled strategic cross-trading, Equations (2) and (3) represent a noisy rational expectations equilibrium.

2.1.b. *Testable Implications*

Proposition 1 generates unambiguous predictions on direct ($\Lambda(n,n)$) and cross-price ($\Lambda(n,j)$) impact. In the model of Section 2.1, speculators are risk neutral, financially unconstrained, and formulate "fundamentally correct" inference from their private signals ($\frac{\partial \delta_m(j)}{\partial S_{vm}(n)} = 0$ if $\Sigma_v(n,j) = 0$), while noise trading is uncorrelated across assets (Σ_z is diagonal). Hence, neither correlated information shocks (King and Wadhwani, 1990; Chan, 1993), correlated liquidity shocks (Bernhardt and Taub, 2008; see also Section 5.2), nor portfolio rebalancing (Kodres and Pritsker, 2002) drive their cross-trading decisions. Nonetheless, Proposition 1 implies that if the underlying economy is *fundamentally interconnected*—a nondiagonal Σ_v—the equilibrium market liquidity matrix Λ of Equation (4) is also nondiagonal: Order flow in one security has a contemporaneous impact on the equilibrium prices of many securities ($\Lambda(n,j) \neq 0$)—even those whose terminal values are unrelated to that security's payoff ($\Sigma_v(n,j) = 0$). Such an impact reflects both (1) speculators' strategic trading activity to affect the MMs' inference from the observed order flow; and (2) the MMs' attempt to learn from it about the traded assets' payoffs v as well as to be compensated for the losses they anticipate from it by their expected profits from noise trading.

Remark 1.
If the economy is fundamentally interconnected there exist cross-price impact, even among fundamentally unrelated assets.

The number of speculators (M) and the correlation among their private information (ρ) affect both direct and cross-price impact. The intensity of competition among speculators influences their ability to attenuate the informativeness of the order flow in each security. More numerous speculators trade more aggressively—that is, their aggregate amount of trading is higher—in every asset because competition among them precludes any collusive trading strategy.[3] This behavior reduces the intensity of adverse selection for the MMs in each market, thus leading to lower direct and absolute (i.e., unsigned) cross-price impact (lower $\Lambda(n,n)$ and $|\Lambda(n,j)|$).

The heterogeneity of speculators' signals moderates their trading aggressiveness. When information is less correlated (smaller ρ), each speculator has some monopoly power on her signal vector, because at least part of it is known exclusively to her. Hence, they trade more cautiously—that is, their absolute amount of trading is lower—in each asset to reveal less of their own information advantage δ_m.[4] This "quasi-monopolistic" behavior makes the MMs more vulnerable to adverse selection. However, the smaller is ρ the lower is the precision of each speculator's private signal of v (as $\Sigma_s = \frac{1}{\rho}\Sigma_v$), hence the less severe is adverse selection for the MMs in all assets. In the presence of many, thus less cautious, speculators (i.e., for nontrivial M, as common to most financial markets), the former effect dominates the latter and both $\Lambda(n,n)$ and $|\Lambda(n,j)|$ increase for lower ρ. Corollary 1 summarizes these empirical implications of our model.

Corollary 1.
Under most parametrizations, direct and absolute cross-price impact are decreasing in the number of speculators and increasing in the heterogeneity of their information.

To gain further insight into these results, we construct a simple numerical example along the lines of Pasquariello (2007). We assume that there are three assets in the economy ($N=3$), that their liquidation values are related to each other by way of the following parametrization of Σ_v:

$$\Sigma_v = \begin{bmatrix} 2 & 0.5 & 0 \\ 0.5 & 1.5 & 0.5 \\ 0 & 0.5 & 2 \end{bmatrix}, \tag{5}$$

and that $\sigma_z^2 = 1$. According to Equation (5), assets 1 and 3 are fundamentally unrelated (i.e., $cov[v(1), v(3)] = 0$) yet both exposed to asset 2 ($cov[v(1), v(2)] > 0$ and $cov[v(2), v(3)] > 0$). We then focus on asset 1 and

[3] For instance, in the limit, if M speculators were *homogeneously* informed—that is, if $\rho = 1$ such that $\Sigma_s = \Sigma_v$, $S_{vm} = v$, and $X_m = \frac{\sigma_z}{\sqrt{M}}\Sigma_v^{-1/2}(v - P_0)$—it can be shown that the finite difference $\Delta|MX_m| = |(M+1)X_m(\text{at}M+1)| - |MX_m(\text{at}M)| = \frac{\sigma_z}{\sqrt{\rho}}\left(\sqrt{M+1} - \sqrt{M}\right)\left|\Sigma_v^{-1/2}(v - P_0)\right| > \underline{0}$.

[4] In particular, $\frac{\partial|X_m|}{\partial\rho} = \frac{\sigma_z}{2\sqrt{M\rho}}\left|\Sigma_v^{-1/2}(S_{vm} - P_0)\right| > \underline{0}$.

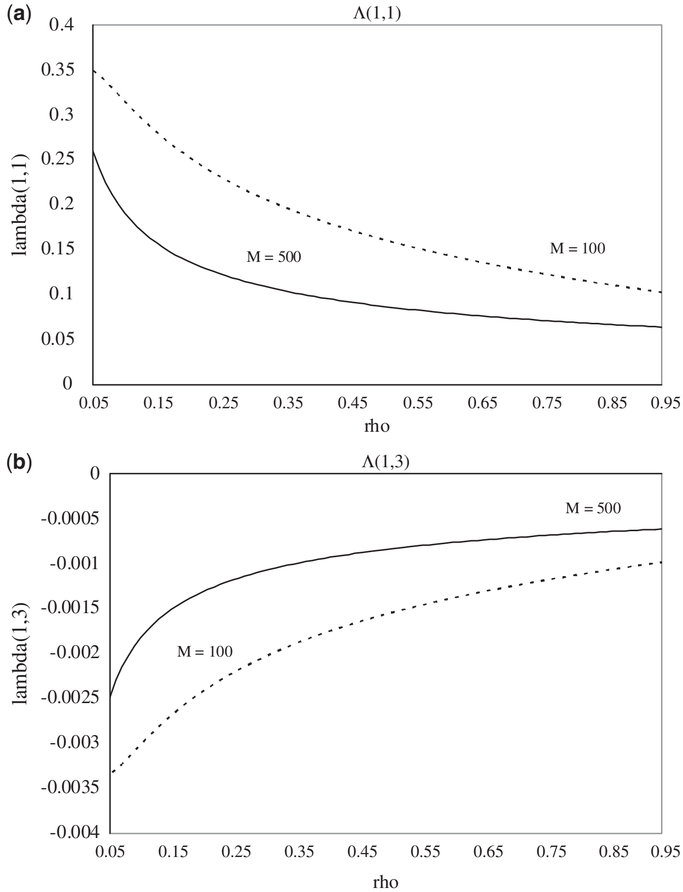

Figure 1. Three-asset economy: measures of liquidity. This figure plots measures of direct ($\Lambda(1,1)$) and cross-price impact ($\Lambda(1,3)$) for the three-asset economy ($N=3$) parametrized in Section 2.1.b as a function of the degree of information heterogeneity among speculators (ρ) in the presence of M of them. Specifically, we plot $\Lambda(1,1)$ (a), and $\Lambda(1,3)$ (b), as a function of $\rho \in [0.05, 0.95]$ when Σ_v is given by Equation (5), $\sigma_z^2 = 1$, and $M = 100$ (dotted line) or $M=500$ (solid line).

plot its equilibrium direct ($\Lambda(1,1)$) and cross-price impact from asset 3 ($\Lambda(1,3)$) as a function of the private signal correlation parameter ρ—in Figures 1a and 1b, respectively—for $M=100$ (dotted line) and $M=500$ (solid line).

As a result of speculators' strategic cross-trading and MMs' cross-inference, order flow in asset 3 impacts the equilibrium price of asset 1, although their terminal payoffs are unrelated: $\Lambda(1,3) \neq 0$ in Figure 1b although

$cov[v(1), v(3)] = 0$. For instance, ceteris paribus, a negative private information shock to asset 1 alone (i.e., to $\delta_m(1)$ alone) prompts imperfectly competitive speculators, moving first and aware of MMs' potential cross-inference, not only to sell asset 1 $\left(\frac{\partial X_m(1)}{\partial \delta_m(1)} > 0, \text{as expected}\right)$ but also to buy asset 2 $\left(\frac{\partial X_m(2)}{\partial \delta_m(1)} < 0\right)$ and to sell asset 3 $\left(\frac{\partial X_m(3)}{\partial \delta_m(1)} > 0\right)$. The latter two trades are to minimize the dissipation of private information and profits that would occur if speculators sold asset 1 alone. Intuitively, selling asset 1 alone leads the MMs to a steeper downward revision of its price (via $\Lambda(1,1) > 0$), hence to lower expected profits for speculators. In Kyle (1985), this possibility induces speculators to more cautious selling of asset 1. In the economy of Equation (5), speculators can also engage in strategic trading in other assets. Speculators' purchases of asset 2 raise the possibility that a positive shock to the common portion of the payoffs of both assets 1 and 2 may have occurred (as $cov[v(1), v(2)] > 0$). Thus, these purchases may *attenuate* the MMs' downward revision of the price of asset 1, yet at the cost of a potential upward revision of the price of asset 2. Speculators' sales of asset 3 raise the possibility that a negative shock to the payoffs of both assets 2 and 3 may have occurred (as $cov[v(2), v(3)] > 0$). Thus, these sales may attenuate the MMs' potential upward revision of the price of asset 2, yet at the cost of a potential downward revision of the price of asset 3. Sales of asset 3 also raise the possibility that the accompanying purchases of asset 2 may stem from positive common information about $v(1)$ and $v(2)$ (rather than about $v(2)$ and $v(3)$). Thus, these sales may attenuate the MMs' downward revision of the price of asset 1 as well.

Aware of this potential strategic cross-trading activity, the MMs make the equilibrium price of asset 1 ($P_1(1)$) sensitive to observed order flow not only in asset 1 ($\Lambda(1,1) > 0$) but also in assets 2 ($\Lambda(1,2) > 0$) and 3 ($\Lambda(1,3) < 0$). Intuitively, a positive $\Lambda(1,2)$ allows the MMs to account for potentially positive common information about $v(1)$ and $v(2)$ from purchases in asset 2 (as $cov[v(1), v(2)] > 0$) when setting $P_1(1)$. Besides, a negative $\Lambda(1,3)$ allows the MMs to account for further potentially positive common information about $v(1)$ and $v(2)$ from purchases in asset 2—in light of potentially negative common information about $v(2)$ and $v(3)$ from sales of asset 3 (as $cov[v(2), v(3)] > 0$)—when setting $P_1(1)$. Figure 1 also shows that a smaller number of speculators (lower M) or greater information heterogeneity among them (lower ρ) intensify such trading activity, and thus worsening MMs' adverse selection problems and increasing both direct ($\Lambda(1,1)$) and absolute cross-price impact ($\Lambda(1,3)$) for asset 1.

3. Data Description

We test the implications of the model of Section 2 using a comprehensive sample of U.S. stock market transaction-level data, and firm-level characteristics.

3.1 U.S. STOCK MARKET DATA

We use intraday, transaction-level TAQ data (trades and quotes) during regular market hours (9:30 a.m. to 4 p.m. ET) for all domestic common stocks (CRSP share code 10 or 11) listed on the NYSE and the NASDAQ between January 1 1993 and June 30 2004 (2,889 trading days). We exclude Real Estate Investment Trusts, closed-end funds, foreign stocks, and American Depository Receipts since their trading characteristics might differ from those of ordinary equities (Chordia and Subrahmanyam, 2004; Boehmer and Wu, 2008). Firm-level accounting information (e.g., quarterly earnings-per-share (EPS)) is from the COMPUSTAT database. Merging TAQ, CRSP, and COMPUSTAT data yields a sample of 3,773 firms (unique identifiers) over our sample period.

We filter the TAQ data by deleting a small number of trades and quotes representing possible data error (e.g., negative prices or quoted depths) or with unusual characteristics (as listed in Bessembinder, 1999, footnote 5). Consistent with the vast literature using TAQ data (Hasbrouck, 2007), we then sign intraday trades using the Lee and Ready (1991) algorithm: (1) If a transaction occurs above (below) the prevailing quote mid-point, we label it a purchase (sale); (2) if a transaction occurs at the quote mid-point, we label it a purchase (sale) if the sign of the last price change is positive (negative). Assigning the direction of trades via the Hasbrouck (1988, 1991) algorithm leads to qualitatively and quantitatively similar inference. As in Bessembinder (2003a), we do not allow for a five-second lag between TAQ reports and compare exchange quotes from NYSE (NASDAQ) exclusively with NYSE (NASDAQ) transaction prices—that is, we only consider order flow taking place in the listing exchange—since off-exchange quotations (e.g., from regional stock exchanges) rarely improve on the exchange quote (Blume and Goldstein, 1997).

Our model, a multi-asset extension of Kyle (1985), conjectures a relationship between a firm's stock price changes and both its own and other firms' net order flow. Chordia and Subrahmanyam (2004, p. 486) observe that "the Kyle setting is more naturally applicable in the context of signed order imbalances over a time interval, as opposed to trade-by-trade data, since the theory is not one of sequential trades by individual traders." Jones, Kaul, and Lipson (1994) and Chordia and Subrahmanyam (2004) also show that the number of transactions has greater explanatory power for stock return fluctuations than dollar trading volume. Accordingly, in this article, we follow Chordia and Subrahmanyam (2004) and Boehmer and Wu (2008), among others, and define the net order flow (i.e., order imbalance) in firm i on day t, $\omega_{i,t}$, as the estimated daily number of buyer-initiated trades ($\mathrm{BUYNUM}_{i,t}$) minus the estimated daily number of seller-initiated trades ($\mathrm{SELLNUM}_{i,t}$) scaled by the total number of trades on day t as follows:

$$\omega_{i,t} = \frac{\mathrm{BUYNUM}_{i,t} - \mathrm{SELLNUM}_{i,t}}{\mathrm{BUYNUM}_{i,t} + \mathrm{SELLNUM}_{i,t}}. \tag{6}$$

We divide the buy–sell imbalance by the total number of trades in Equation (6) to eliminate the impact of total trading activity (Chordia and Subrahmanyam, 2004). In unreported analysis, we find our inference to be robust to defining order imbalance as the net scaled dollar trading volume (Jones, Kaul, and Lipson, 1994) or to using alternative normalizations of the buy–sell imbalance (e.g., by scaling it by the number of shares outstanding or a moving average of the total number of trades over the trailing year).

3.2 INFORMATION VARIABLES

According to the model of Section 2, the magnitude of equilibrium cross-price impact among traded assets depends on the extent of market-wide information heterogeneity among speculators (ρ) and on the number of speculators in the economy (M).

We use professional forecasts of either individual firm's long-term EPS growth or U.S. macroeconomic announcements to proxy for the beliefs of sophisticated market participants about U.S. stocks' fundamentals. The standard deviation across professional forecasts is a commonly used measure of aggregate and stock-level information heterogeneity (Diether, Malloy, and Scherbina, 2002; Kallberg and Pasquariello, 2008).

We obtain our first proxy for ρ by using the unadjusted I/B/E/S Summary History database of analyst forecasts of the long-term growth of individual stocks' EPS. Long-term growth forecasts are less likely to be biased by firms' potential "earnings guidance" (Yu, 2011) and normalization for cross-firm comparability (Qu, Starks, and Yan, 2004). The inference that follows is robust to use fiscal-year EPS forecasts. We define the diversity of opinion about the long-term prospects of each firm i in the TAQ/CRSP/COMPUSTAT sample in each month m between January 1993 and June 2004 as the standard deviation across multiple (i.e., two or more) analyst forecasts of that firm's long-term EPS growth (when available),

$\mathrm{SDLTEPS}_{i,m}$.[5] Following Kallberg and Pasquariello (2008) and Yu (2011), we then compute our measure of market-wide information heterogeneity in month m, $\mathrm{SDLTEPS}_m$ (Figure 2a), as a simple average of firm-level dispersion of opinion in that month. Equal-weighting adjusts for the relatively poor coverage of small stocks in our merged TAQ/CRSP/COMPUSTAT dataset. We discuss this issue in greater detail in Section 4. However, our inference is insensitive to computing $\mathrm{SDLTEPS}_m$ as a value-weighted average of firm-level forecast dispersions. Yu (2011) shows that this measure successfully captures the common component of differences in investors' opinions about the future prospects of individual stocks in the U.S. equity market.

Our second proxy for ρ is based upon the professional forecasts of 18 U.S. macroeconomic announcements from the International Money Market Services Inc. (MMS) real-time database. These data are available to us between January 1993 and December 2000.[6] As in Pasquariello and Vega (2007, 2009), we define the dispersion of beliefs among speculators about a macroeconomic announcement p in month m as the standard deviation of its forecasts in that month, $\mathrm{SDMMS}_{p,m}$. We then compute the aggregate degree of information heterogeneity about macroeconomic fundamentals in month m, SDMMS_m (Figure 2b), as a simple average of all standardized announcement-level forecast dispersions in that month. Specifically, on each month m of the sample period, we divide the difference between the standard deviation of professional forecasts for each macroeconomic announcement p in that month (from MMS) and its sample mean by its sample standard deviation (see Pasquariello and Vega, 2007; Equation (7)). Normalization is necessary because units of measurement differ across announcements. We also shift the mean of SDMMS_m by a factor of 10 to ensure that SDMMS_m is always positive.

We use three different proxies for the number of informed traders in the U.S. stock market: average analyst coverage across firms in month; the average fraction of shares held by institutional investors; and the per-share intensity of informed trading. According to prior literature (Brennan and Subrahmanyam, 1995; Chordia, Huh, and Subrahmanyam, 2007), analyst

[5] Diether, Malloy, and Scherbina (2002) describe similarities and differences between the I/B/E/S Summary History and Detailed History databases. In unreported analysis, we find analogous results when obtaining analyst forecast data from the latter.

[6] Since being acquired by Informa in 2003, MMS discontinued its survey services. The 18 announcements are: GDP Advance, GDP Preliminary, GPD Final, Nonfarm Payroll Employment, Retail Sales, Industrial Production, New Home Sales, Durable Goods Orders, Factory Orders, Construction Spending, Trade Balance, Producer Price Index, Consumer Price Index, Consumer Confidence Index, NAPM Index, Housing Starts, Index of Leading Indicators, and Initial Unemployment Claims.

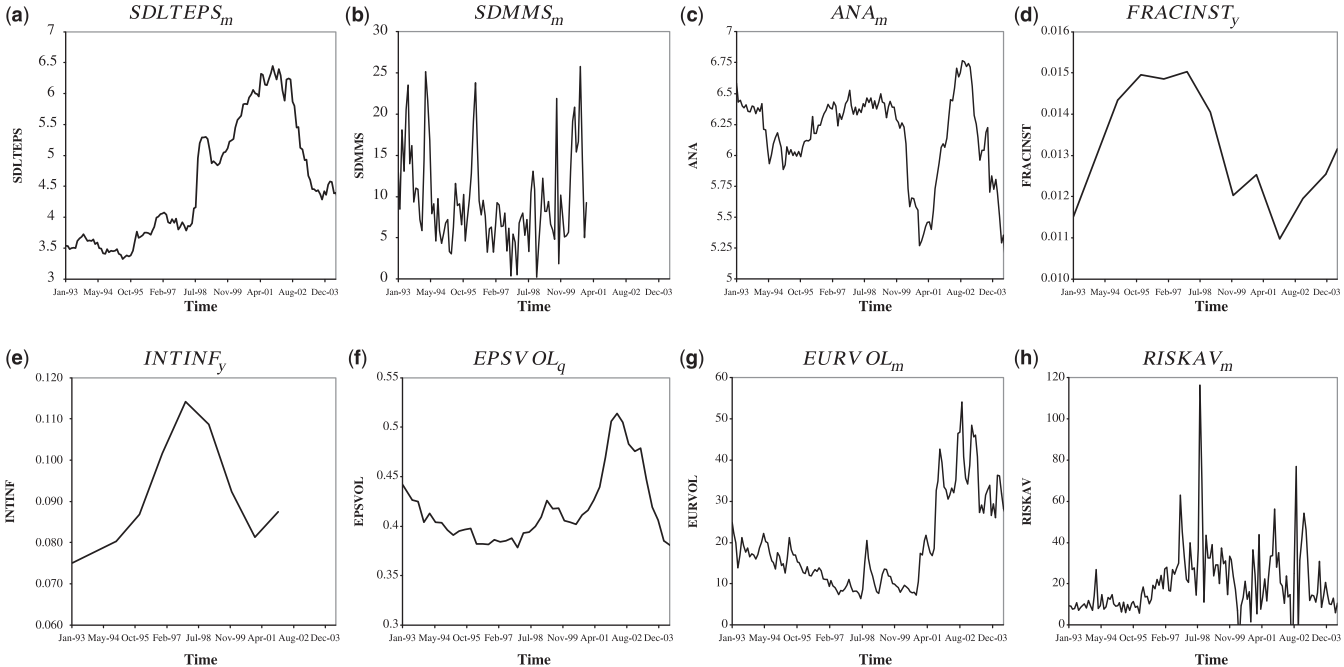

Figure 2. Plots of market-wide aggregates. (a) $\mathrm{SDLTEPS}_m$ plot, the equal-weighted average of firm-level standard deviations of analyst forecasts of long-term EPS growth (in percentage, from I/B/E/S; see Section 3.2) in month m. (b) SDMMS_m plot, the simple average of standardized standard deviation of professional forecasts of 18 U.S. macroeconomic announcements (from MMS; see Section 3.2). (c) ANA_m plot, the equal-weighted average of the number of analysts covering each of the firms in our TAQ sample in month m (from I/B/E/S; see Section 4.2.b). (d) $\mathrm{FRACINST}_y$ plot, the equal-weighted average of the fraction of U.S. shares outstanding held by mutual funds in year y (from Thompson/CDA/Spectrum; see Section 4.2.b). (e) INTINF_y plot, the equal-weighted average of the per-share intensity of informed trading in the U.S. stock market in year y, as estimated by Vega (2006) from the *PIN* model of Easley and O'Hara (1992) (see Section 4.2.b). (f) EPSVOL_q plot, the equal-weighted average of firm-level EPS volatility in calendar quarter q (in dollars, over the most recent eight quarters, from COMPUSTAT; see Section 5.1). (g) EURVOL_m plot, the monthly average of daily Eurodollar implied volatility (in percentage, from Bloomberg; see Section 5.1). (h) RISKAV_m plot, the monthly difference between the end-of-month VIX index of implied volatility of S&P500 options with 30-day fixed maturity and the realized volatility of intraday S&P500 returns over that month (from Bollerslev, Tauchen, and Zhou, 2009; see Section 5.1). All variables are plotted over the interval January 1993 to June 2004, except SDMMS_m (available until December 2000) and INTINF_y (available until December 2001).

coverage relates to informed trading because analysts produce information about the value of a firm that is not publicly accessible to all market participants, but only to a subset of them—sophisticated speculators, or traders with superior analytical ability—and often for a fee. We compute our measure of market-wide number of speculators M as ANA_m (Figure 2c), the equal-weighted average of the number of analysts covering each firm i's long-term EPS growth in our sample and reporting their forecasts of that firm's long-term EPS growth to the I/B/E/S database in month m, $\mathrm{ANA}_{i,m}$. We use averages to adjust for the time-varying number of firms in our sample. We obtain similar inference when replacing ANA_m with either the sum or the value-weighted average of firm-level $\mathrm{ANA}_{i,m}$.

An alternative proxy for M is the fraction of U.S. shares held by institutional investors. Prior literature suggests that institutions are better informed than other market participants (Grinblatt and Titman, 1989; Daniel *et al.*, 1997; Wermers, 1999). We use the Thomson Reuters Mutual Fund Holdings database (formerly known as CDA/Spectrum) to compute the equal-weighted average fraction of U.S. shares held by institutional investors in year y, $\mathrm{FRACINST}_y$ (Figure 2d). We also measure M as the per-share intensity of informed trading in the U.S. stock market, INTINF_y (Figure 2e). Specifically, we compute INTINF_y as the equal-weighted average of firm-level estimates (at the annual frequency) of the arrival rate of informed trading ($\mu_{i,y}$, from Easley and O'Hara's (1992) PIN model) divided by the corresponding total number of shares outstanding. INTINF_y can only be computed between 1993 and 2001 because the significant increase in the number of transactions in the U.S. stock market in subsequent years has made estimation of the PIN model unfeasible. For further details, see Vega (2006). In contrast to ANA_m, the average *number* of mutual funds holding U.S. stocks and the arrival rate of informed trading (unreported) are nonstationary, having both grown exponentially over our sample period along with the number of shares outstanding. Dividing them by the latter makes the resulting ratios $\mathrm{FRACINST}_y$ and INTINF_y not only stationary but also conceptually closer to the comparative statics within our theoretical model, in which we change the number of speculators while keeping all else fixed (including the number of assets traded).

The plots of our information environment proxies in Figure 2 suggest that, over the sample period 1993–2004, aggregate dispersion of beliefs in the U.S. stock market is large,[7] time-varying, positively correlated across different

[7] For example, market-wide information heterogeneity averages 4.6% when measured by the standard deviation of long-term EPS growth forecasts ($\mathrm{SDLTEPS}_m$), versus an equal-weighted mean of those forecasts of 16.8%.

proxies (albeit not strongly so), and weakly negatively correlated with average analyst coverage. Common disagreement is low in the mid-1990s, sharply increases and declines in correspondence with the Internet stock bubble, and stays historically high afterward. The average number of analysts following a U.S. firm is instead relatively high in the 1990s, and significantly more volatile around the Global Settlement of 2002 (Juergens and Lindsey, 2009); both average institutional ownership and informed trading intensity experience a similarly sharp decrease in the late 1990s. Accordingly, the three proxies for the number of speculators in the market are positively correlated. In particular, $\mathrm{FRACINST}_y$ and INTINF_y have a very high positive correlation (0.97), but are only weakly positively correlated with ANA_m (approximately 0.10). These dynamics are consistent with those reported in recent studies using similar proxies (Chordia, Huh, and Subrahmanyam, 2007; Pasquariello and Vega, 2007; Yu, 2011).

4. Empirical Analysis

The model of Section 2 generates several implications for direct and cross-price impact in the U.S. equity market that we now test in this section.

4.1 CROSS-PRICE IMPACT IN THE U.S. STOCK MARKET

In the context of our model, direct price impact of any stock i is defined as the marginal contemporaneous impact of trading in stock i on its equilibrium price, $\lambda_{ii,0}$. Similarly, cross-price impact between any two stocks i and h is defined as the marginal contemporaneous impact of trading in stock h on the equilibrium price of stock i, $\lambda_{ih,0}$. Ideally, we would need to estimate direct and cross-price impact for each of the stocks in our sample—that is, the liquidity matrix Λ of Equation (4) for the entire U.S. stock market. This is a challenging task. Measures of direct price impact are typically estimated as the slopes of regressions of stock returns on direct order imbalance over either intraday or daily time intervals (Hasbrouck, 2007). A natural extension of this procedure to our setting would be to assess the sensitivity of the returns of each stock to both its own and each of all other stocks' order imbalances. The large number of stocks in our database and the relative scarcity of trades for some of them makes the literal implementation of this extension impractical. In light of these considerations, we use two complementary approaches by focusing on either industry-level stock portfolios or random stock pairs.

Table 1. Summary statistics

This table reports summary statistics for daily, industry-level, equal-weighted returns, $r_{n,t}$ (as defined in Section 4.1.a), and net order flow, $\omega_{n,t}$ (net scaled number of transactions; see Section 4.1.a), in our merged TAQ/CRSP/COMPUSTAT dataset between January 1993 and June 2004 (2,889 observations). Both variables are in percentages, that is, are multiplied by 100. A *, **, or *** indicates significance at the 10%, 5%, or 1% level, respectively.

	Returns $r_{n,t}$ (%)					Net Order Flow $\omega_{n,t}$ (%)				
	Mean	Median	stdev	1%	99%	Mean	Median	stdev	1%	99%
Durables	−0.004	0.029	1.201	−3.136	3.123	3.182***	3.112	8.287	−15.670	21.802
Nondurables	−0.004	0.038	0.839	−2.279	2.130	2.241***	2.148	7.583	−15.467	19.509
Manufacturing	−0.004	0.029	1.105	−2.904	3.071	3.530***	3.498	7.094	−13.407	19.421
Energy	0.023	0.033	1.363	−3.662	3.580	3.867***	4.228	8.687	−17.149	22.057
HighTech	−0.032	0.118	1.936	−5.330	5.052	−0.287**	0.092	7.318	−17.019	15.619
Telecom	−0.044	0.048	1.648	−4.851	4.360	0.685***	0.717	7.127	−15.697	16.607
Shops	−0.025	0.027	1.134	−2.902	2.924	2.273***	2.275	6.810	−13.499	17.890
Health	−0.003	0.065	1.294	−3.586	3.232	0.355***	0.322	6.876	−15.175	16.824
Utilities	−0.005	0.022	0.970	−2.933	2.484	−1.179***	−0.494	11.452	−27.582	21.391
Other	−0.007	0.063	1.114	−3.126	2.940	2.629***	2.636	6.974	−13.485	18.611

4.1.a. *Industry-Level Analysis*

In the first approach, we proceed by (1) aggregating the stocks in our sample into a smaller number of industry portfolios; and (2) estimating direct and cross-price impact among these portfolios. This approach reduces both the dimension of the liquidity matrix Λ to be estimated and the potential impact of idiosyncratic noise on its estimation. We sort the 3,773 firms in our sample into the 10 broad industry groupings proposed by Fama and French (1988): Durables, Nondurables, Manufacturing, Energy, HighTech, Telecom, Shops, Health, Utilities, and Other (from French's research website). We then use daily, firm-level close-to-close mid-point stock returns ($r_{i,t}$, from TAQ) and estimated order imbalance ($\omega_{i,t}$ of Equation (6)) to compute daily, equal-weighted industry-level portfolio returns ($r_{n,t}$) and net order flow ($\omega_{n,t}$), respectively.[8] Chordia and Subrahmanyam (2004) document that contemporaneous price impact of daily order imbalance is increasing in firm size. Thus, using value-weighted net order flow has the potential to favorably bias our analysis in a systematic way. Value-weighted averages lead to similar inference. We report summary statistics for $r_{n,t}$ and $\omega_{n,t}$ in Table I.

[8] Using close-to-close mid-point returns mitigates the bid-ask bounce bias in daily stock returns (e.g., see the discussion in Chordia and Subrahmanyam, 2004; Boehmer and Wu, 2008). Qualitatively similar inference ensues from using CRSP returns or open-to-close mid-point returns.

Mean order imbalance $\omega_{n,t}$ is positive for most industries, suggesting that buying pressure was predominant among market orders over our sample period. Average daily firm-level order imbalance $\omega_{i,t}$ is also positive (about 1.73%) and in line with previous studies (Boehmer and Wu, 2008).

The model of Section 2 implies that speculators' strategic cross-trading leads to equilibrium cross-price impact (even among fundamentally unrelated assets) as long as the underlying economy is fundamentally interconnected (Remark 1). To assess the extent to which the 10 industries listed earlier are fundamentally related over our sample period, we estimate (and report in Table II) a correlation matrix of industrywide, equal-weighted averages of firm-level quarterly earnings in two steps. First, for every firm i in our dataset, we obtain its quarterly EPS ($\mathrm{EPS}_{i,q}$, basic, excluding extraordinary items, for calendar quarter q) over our sample period 1993–2004 (when available). Second, we estimate Spearman correlations ($\rho_{n,j}$) of equal-weighted averages of $\mathrm{EPS}_{i,q}$ within each industry n and quarter q, $\mathrm{EPS}_{n,q}$. Not surprisingly, Table II indicates that the U.S. economy, as represented by its publicly traded companies, is fundamentally interconnected. Importantly, Table II also suggests that cross-industry earnings correlations are not uniformly high, positive, and statistically significant but vary pronouncedly from the highest (0.774 between Manufacturing and HighTech) to the lowest (0.009 between Energy and Manufacturing) in absolute value terms.

Next, we estimate the liquidity matrix among our 10 industry portfolios. The expression for the equilibrium prices of all assets in Proposition 1 (Equation (2)) translates naturally in the following set of 10 regression models:

$$r_{n,t} = \alpha_n + \sum\nolimits_{j=1}^{10} \lambda_{nj,0}\omega_{j,t} + \varepsilon_{n,t}. \tag{7}$$

According to our model, we expect estimates for industry n's direct price impact of its own order imbalance, $\lambda_{nn,0}$, to be positive and estimates for industry n's cross-price impact of other industries' order imbalances, $\lambda_{nj,0}$ for $n \neq j$, to be significant.

Even in the absence of the strategic, informational cross-trading activity described in our model, inventory considerations may induce correlation between industry portfolio returns and cross-industry order flow. For instance, dealers' attempts to manage inventory fluctuations correlated across individual assets—because of market-wide dynamics in cash flows, trading volume, inventory carrying costs, volatility, or risk-bearing capacity (Chordia, Roll, and Subrahmanyam, 2000; Chordia and Subrahmanyam, 2004; Andrade, Chang, and Seasholes, 2008; Comerton-Forde *et al.*, 2010)—may eventually generate cross-price impact even when order

Table II. Industry-level earnings correlations

This table reports estimated Spearman correlations (EPS CORR, over the sample period January 1993 to June 2004, 46 observations) among industry-level, equal-weighted averages of quarterly earnings (EPS basic, excluding extraordinary items) of the corresponding firms (as defined in Section 4.1.a) in each of the 10 industries in our sample (Durables, Nondurables, Manufacturing, Energy, HighTech, Telecom, Shops, Health, Utilities, and Other). A *, **, or *** indicates significance at the 10%, 5%, or 1% level, respectively.

	EPS CORR									
	Durables	Nondurables	Manufacturing	Energy	HighTech	Telecom	Shops	Health	Utilities	Other
Durables	**1**									
Nondurables	0.026	**1**								
Manufacturing	0.577***	0.246*	**1**							
Energy	−0.079	0.049	0.009	**1**						
HighTech	0.395***	0.104	0.784***	−0.185	**1**					
Telecom	0.128	0.152	0.505***	−0.118	0.763***	**1**				
Shops	0.045	−0.021	0.304**	−0.240	0.583***	0.580***	**1**			
Health	0.235	0.107	0.537***	−0.012	0.658***	0.742***	0.565***	**1**		
Utilities	−0.035	0.059	0.266	0.394***	0.139	0.021	−0.047	0.188	**1**	
Other	0.423***	0.219	0.731***	−0.013	0.782***	0.660***	0.423***	0.593***	0.090	**1**

imbalances have no information content. To control for this possibility, we include lagged values of direct and cross-industry order imbalances in Equation (7), in the spirit of Hasbrouck (1991) and Chordia and Subrahmanyam (2004). Hasbrouck (1991) argues that trades in an asset have a persistent direct impact on its prices if due to information shocks, but a transitory impact if due to noninformational (e.g., liquidity or inventory-driven) shocks and other microstructure imperfections (price discreteness, bid-ask bounce, exchange-mandated price smoothing, or order fragmentation; see also Hasbrouck, 2002). Consistently, we conjecture that cross-price impact in Equation (7) may be attributed to informational cross-trading and cross-inference (as advocated by our model) if it is persistent—that is, if it is robust to transient, lagged noninformational effects.

Noninformational commonality in prices and trading activity may also lead to cross-price impact, even in the absence of strategic, informational cross-trading. For instance, Chordia, Roll, and Subrahmanyam (2000) observe that market-wide trading activity may be sensitive to general swings in stock prices. Hasbrouck and Seppi (2001) suggest that their evidence of common factors in prices and order flow of Dow 30 stocks may be attributed to such market-wide liquidity shocks as portfolio substitution. Barberis, Shleifer, and Wurgler (2005) argue that investors' portfolio rebalancing activity may be triggered by noninformational shifts in the composition of broad categories and indexes (category and index investing) or other fixed subsets of all available securities (habitat investing). Alternatively, market-wide information shocks may also cause correlated trading and price changes (King and Wadhwani, 1990; Chan, 1993; Hasbrouck and Seppi, 2001). For example, Chan (1993) develops a multi-asset model populated only by privately informed MMs who by construction are not allowed to trade across stocks. In that setting, Chan (1993) shows that the presence of a market-wide information component in stock payoffs leads to cross-autocorrelation in equilibrium stock price changes because the price of a stock may be informative about the payoff of another stock. In Section 5.1, we examine in greater detail the potential impact of these and other theories of cross-trading activity within the U.S. equity market on our inference. At this preliminary stage, we control for the extent of portfolio rebalancing and correlated, market-wide informational trading by including daily equal-weighted returns on all NYSE and NASDAQ stocks in our sample ($r_{M,t}$) in Equation (7).[9]

[9] The inclusion of market returns also allows to reduce potential cross-correlations in error terms among industries (e.g., Chordia and Subrahmanyam, 2004). The estimated coefficients on $r_{M,t}$, not reported here, are in line with those in the literature for similar

Equation (7) is based on the presumption—rooted in the microstructure literature—of a causal link from trades to price changes. As such, we further amend it to include lagged returns of all industry portfolios in our sample to control for (lagged adjustment to) any public direct and cross-industry information already set in those portfolios' recent price change history, in the spirit of Hasbrouck (1991), Chan (1993), and Chordia, Sarkar, and Subrahmanyam (2011). The inclusion of direct and cross-autocorrelation terms in Equation (7) only weakens our evidence. The ensuing regression model for the estimation of direct and cross-industry price impact in the U.S. equity market is given by:

$$r_{n,t} = \alpha_n + \beta_n r_{M,t} + \sum\nolimits_{j=1}^{10} \sum\nolimits_{l=1}^{L} \gamma_{nj,l} r_{j,t-l} + \sum\nolimits_{j=1}^{10} \sum\nolimits_{l=0}^{L} \lambda_{nj,l} \omega_{j,t-l} + \varepsilon_{n,t}, \tag{8}$$

where $L = 3$ trading days and $\sum_{l=0}^{L} \lambda_{nn,l}$ and $\sum_{l=0}^{L} \lambda_{nj,l}$ are measures of cumulative direct and cross-price impact of order flow, respectively. We estimate Equation (8) for each of the 10 industries separately by ordinary least squares (OLS)—further correcting the standard errors for heteroskedasticity and serial correlation—and report the resulting estimates of cumulative direct ($\sum_{l=0}^{L} \lambda_{nn,l}$, on the diagonal) and cross-industry price impact ($\sum_{l=0}^{L} \lambda_{nj,l}$, off the diagonal) in the corresponding rows of Table III.[10]

Consistent with the literature, we interpret statistically significant estimates of these measures as evidence of persistent, informational (direct or cross-industry) price impact, that is, price impact that persists up to three trading days and so is more likely to be attributable to informational trading than to inventory management, marketwide trading, and other noninformational effects. It is worth emphasizing the conservative nature

industry portfolios (e.g., Bernanke and Kuttner, 2005). We obtain qualitatively similar results when omitting $r_{M,t}$, when including daily value-weighted market returns, when using market-adjusted industry returns ($r_{n,t} - r_{M,t}$), or when replacing $r_{M,t}$ with the three Fama-French factors (market excess returns MKT_t, size SMB_t, and book-to-market HML_t; Fama and French, 1993)—available on French's website—to control for both a broader set of systematic sources of risk and other popular forms of category investing (e.g., small-cap versus large cap, or value versus growth).

[10]Joint estimation of Equation (8) by Feasible Generalized Least Squares (FGLS) leads to the same, efficient coefficient estimates since the resulting ten 10 staked regression models have identical explanatory variables (e.g., Greene, 1997, p. 676). Unreported analysis indicates that the corresponding adjusted R^2 (R_a^2) are higher than when replacing industry-level, aggregate net scaled number of transactions ($\omega_{n,t}$) in Equation (8) with industry-level, aggregate net scaled trading volume, in line with Jones, Kaul, and Lipson (1994) and Chordia and Subrahmanyam (2004).

Table III. Direct and cross-industry price impact

This table reports estimates (IMPACT, in column) of persistent direct and cross-industry price impact ($\sum_{l=0}^{L} \lambda_{nn,l}$, on the diagonal, and $\sum_{l=0}^{L} \lambda_{nj,l}$, off the diagonal, respectively) for each industry portfolio (in row),from the regression model of Equation (8):

$$r_{n,t} = \alpha_n + \beta_n r_{M,t} + \sum_{j=1}^{10} \sum_{l=1}^{L} \gamma_{nj,l} r_{j,t-l} + \sum_{j=1}^{10} \sum_{l=0}^{L} \lambda_{nj,l} \omega_{j,t-l} + \varepsilon_{n,t},$$

where $r_{n,t-l}$ is the equal-weighted average of daily stock returns in industry portfolio n on day $t-l$, $r_{M,t}$ is the equal-weighted daily return on all NYSE and NASDAQ stocks, $\omega_{n,t-l}$ is the equal-weighted, industry-level net order flow (net scaled number of transactions) on day $t-l$, and $L=3$. We estimate Equation (8) by OLS over the sample period January 1993 to June 2004 (2,889 observations) and assess the statistical significance of the estimated coefficients with Newey–West standard errors. Coefficient estimates are multiplied by 100. R_a^2 is the adjusted R^2. A *, **, or *** indicates significance at the 10%, 5%, or 1% level, respectively.

	IMPACT										
	Durables	Nondurables	Manufacturing	Energy	HighTech	Telecom	Shops	Health	Utilities	Other	R_a^2
Durables	**2.166*****	−2.306***	1.769***	0.392	−1.610***	0.276	−0.029	0.852*	−0.149	−0.729	69%
Nondurables	0.339	**2.556*****	−0.369	−0.439*	−0.436	−0.297	0.313	−0.318	−0.388**	−0.684	67%
Manufacturing	1.047***	−0.833*	**3.324*****	−0.102	−0.868**	0.154	−0.140	−0.520**	−0.343**	−1.533***	84%
Energy	0.734	−2.749***	−1.447*	**5.985*****	3.773***	0.109	−1.542	−0.625	0.451	−2.296*	54%
HighTech	0.153	1.252***	−0.696	0.115	**3.053*****	−0.293	0.001	−1.028***	0.178	−2.478***	93%
Telecom	−0.982**	1.327**	1.250*	−1.042***	−1.170**	**3.855*****	−0.545	−0.458	−0.560**	−2.109**	81%
Shops	−0.625**	−0.157	−0.038	−0.141	−1.984***	−0.386	**5.803*****	0.159	−0.567***	−1.520***	84%
Health	0.512	−2.611***	−0.242	−0.021	−2.581***	0.530	0.040	**4.670*****	0.110	0.046	81%
Utilities	−0.405	−1.531**	0.050	0.617*	1.622***	−1.393***	−1.222	−0.197	**1.549*****	0.563	43%
Other	−0.029	−0.245	−0.666**	−0.478***	−1.462***	−0.572**	−1.379***	−0.286	−0.090	**4.390*****	90%

of this approach, since those effects are likely to dissipate within a single trading day (Hasbrouck, 1991, 2007). Estimates of contemporaneous direct and cross-price impact coefficients ($\lambda_{nn,0}$ and $\lambda_{nj,0}$, available on request) yield much stronger support for our model than the cumulative ones reported in the article. Therefore, our inference is based upon an extremely conservative scenario for testing our model.

Table III provides evidence of direct and cross-industry informational effects of industry order flow on daily industry portfolio returns. Consistent with both our model and previous studies (Chordia and Subrahmanyam, 2004; Boehmer and Wu, 2008), estimates of informational direct price impact of net order flow $\sum_{l=0}^{L} \lambda_{nn,l}$ are economically and statistically significant for each of the 10 industries in our sample. For example, a 1 standard deviation shock to an industry's daily direct order imbalance ($\sigma(\omega_{n,t})$, in Table I) affects its own daily, market-adjusted returns by an average of 28.3 basis points ($\sigma(\omega_{n,t}) \sum_{l=0}^{L} \lambda_{nn,l}$). Importantly, and consistent with our model, nearly half of the estimates of informational cross-industry price impact of order imbalance are economically and statistically significant as well—averaging 10.2 basis points per 1 standard deviation shocks to other industries' order flow $\left(\sigma(\omega_{j,t}) \left| \sum_{l=0}^{L} \lambda_{nj,l} \right|\right)$—even among less closely related industries (e.g., among industries displaying low earnings correlations in Table II). For instance, daily Energy returns increase by an average of 27.6 basis points in correspondence with a 1 standard deviation shock to daily order imbalance in HighTech stocks (3.773×7.318)—versus an average of 52 basis points in correspondence with a similar shock to its own order flow (8.687×5.985)—although the historical correlation between quarterly earnings of these two industries is low and statistically insignificant (–0.185, in Table II). As noted earlier, our evidence is stronger if we focus on contemporaneous cross-price impact of order flow. For example, a larger fraction (58%) of (unreported) contemporaneous cross-industry price impact coefficients ($\lambda_{nj,0}$ in Equation (8)) is statistically and economically significant.

Interestingly, many of the statistically significant cross-industry price impact coefficients $\sum_{l=0}^{L} \lambda_{nj,l}$ in Table III are *negative*. This is also consistent with our model, but may not be so with alternative theories of cross-trading activity in the literature that more directly link price impact to the covariance structure of asset fundamentals (Kodres and Pritsker, 2002; Bernhardt and Taub, 2008). In our model, as the numerical example in Section 2.1.b illustrates, strategic cross-trading within an interconnected economy—e.g., the payoff covariance matrix Σ_v of Equation (5), where most (but not all) nondiagonal elements are positive (not unlike the correlation matrix of industry EPS in Table II)—suffices to generate a nontrivial fraction (33%)

of negative cross-price impact in the equilibrium liquidity matrix Λ of Equation (4).

4.1.b. *Firm-Level Analysis*

The evidence in Table III supports the notion, implied by our model, that there is persistent, informational cross-price impact at the level of industry groupings of U.S. stocks. In the second approach, we provide further evidence of cross-price impact in the U.S. equity market at the level on the individual stocks, where noninformational and market-wide informational commonality in prices and trading activity is likely to be lower than for industry portfolios. We do so manageably by repeatedly estimating direct and cross-price impact for any two randomly selected stocks in our sample.

Specifically, we first identify all the stocks in our initial TAQ/CRSP/COMPUSTAT sample with at least nine quarters of common history of quarterly earnings. This requirement, which yields 472 firms, ensures that the correlation of the EPS of any two randomly selected firms, a proxy for the intensity of their fundamental relationship, is computed over a sufficiently long window (at least 2 years). In unreported analysis, we find that the firms in this subsample are larger and tilt less toward growth or HighTech than those in the TAQ/CRSP/COMPUSTAT universe, but are otherwise similarly distributed across industries. For any two stocks i and h randomly drawn from this subset we then compute the correlation of their earnings ($\mathrm{EPS}_{i,q}$ and $\mathrm{EPS}_{h,q}$) and estimate the following version of Equation (8):

$$\begin{aligned} r_{i,t} = \alpha_i + \beta_i r_{M,t} + \sum\nolimits_{l=1}^{L} \gamma_{ii,l} r_{i,t-l} + \sum\nolimits_{l=1}^{L} \gamma_{ih,l} r_{h,t-l} \\ + \sum\nolimits_{l=0}^{L} \lambda_{ii,l}\omega_{i,t-l} + \sum\nolimits_{l=0}^{L} \lambda_{ih,l}\omega_{h,t-l} + \varepsilon_{i,t}, \end{aligned} \tag{9}$$

again by OLS over each stock pair's longest common trading history in TAQ within our sample period. We repeat this procedure 2,000 times, and then compute averages of the ensuing 2,000 estimates of cumulative direct ($\sum_{l=0}^{L} \lambda_{ii,l}$) and pairwise absolute cross-price impact $\left(\left|\sum_{l=0}^{L} \lambda_{ih,l}\right|\right)$, to prevent the aforementioned signed effects from canceling out) on stock i's return ($r_{i,t}$) from a 1 standard deviation shock to its own ($\sigma(\omega_{i,t})$) or the other stock's order imbalance ($\sigma(\omega_{h,t})$), respectively. We report these averages in basis points in Table IV (Panel A), together with averages of those effects within each quintile of stock pairs sorted according to their absolute earnings correlations ($|\rho_{i,h}|$) from the lowest to the highest, as well as averages of those effects when statistically significant.

A priori, we expect the estimation of Equation (9) to yield weaker results than the estimation of Equation (8), because standard errors of regressions of individual stock returns are larger than those of regressions of portfolio stock returns (where idiosyncratic noise may cancel out). Nevertheless, Panel A of Table IV indicates that, consistent with our model (and Table III), cross-price impact among randomly selected stocks is large and statistically significant more often than if due to chance (i.e., to statistical Type I error)—in 14% of the random stock pairs at the 10% level. This evidence is unlikely to be due to model misspecification or omitted return commonality. Our basic empirical strategy (based on Hasbrouck, 1991; Chordia and Subrahmanyam, 2004) is popular in the literature for its robustness. In addition, by virtue of portfolio aggregation, random stock pairs are less likely to display economic relatedness than industry portfolios. For example, over our sample period 1993–2004, the average correlation among the returns of our 2,000 random stock pairs is 0.21, much lower than the average correlation of our 10 industry portfolios, 0.68. Unreported analysis further shows that (1) nearly all (99%) of the contemporaneous direct price impact coefficients for the random stocks in Table IV—$\lambda_{ii,0}$ in Equation (9)—are statistically significant at the 10% level; and (2) as noted earlier, a larger fraction (25%) of the corresponding contemporaneous cross-price impact coefficients among the random stock pairs in Table IV ($\lambda_{ih,0}$ in Equation (9)) are statistically significant at the 10% level. Lower significance levels yield qualitatively similar inference.

Our model also postulates that strategic cross-trading activity may lead to equilibrium cross-price impact even among fundamentally unrelated assets. Accordingly, Table IV reports evidence of significant persistent (i.e., informational) cross-price impact coefficients even within the quintiles of firm pairs displaying low average absolute earnings correlations. For instance, the daily returns of a randomly selected stock within the first such quintile of firm pairs (whose mean $|\rho_{i,h}|$ is 0.04) move by an average of 41.1 basis points from a 1 standard deviation shock to its own order flow—in the 93% of the cases in which such direct impact is statistically significant—and by 13.8 basis points from a 1 standard deviation shock to the order flow of another randomly selected stock—in the 16% of the cases in which such cross-impact is statistically significant. In addition, again consistent with our model (and Table III), 57% of the statistically significant estimates of cumulative cross-price impact in Panel A of Table IV ($\sum_{l=0}^{L} \lambda_{ih,l}$) are negative. Finally, unreported t-tests always strongly reject (at less than the 1% level) the null hypothesis that the average of these estimates within all and each quintile of random stock pairs is zero.

Table IV. Direct and absolute cross-price impact for random stock pairs

This table reports estimates of average persistent direct and absolute cross-stock price impact (DIRECT IMPACT as $\sum_{l=0}^{L} \lambda_{ii,l}$ and CROSS IMPACT as $\left|\sum_{l=0}^{L} \lambda_{ih,l}\right|$) accompanying a 1 standard deviation shock to the corresponding order flow ($\sigma(\omega_{i,t})$ and $\sigma(\omega_{h,t})$) in basis points (i.e., multiplied by 10,000), from the regression model of Equation (9):

$$r_{i,t} = \alpha_i + \beta_i r_{M,t} + \sum_{l=1}^{L} \gamma_{ii,l} r_{i,t-l} + \sum_{l=1}^{L} \gamma_{ih,l} r_{h,t-l} + \sum_{l=0}^{L} \lambda_{ii,l} \omega_{i,t-l} + \sum_{l=0}^{L} \lambda_{ih,l} \omega_{h,t-l} + \varepsilon_{i,t},$$

where $r_{i,t-l}$ is the daily returns of randomly selected stock i on day $t-l$, $r_{M,t}$ is the equal-weighted daily return on all NYSE and NASDAQ stocks, $\omega_{i,t-l}$ is the daily net order flow (net scaled number of transactions) in firm i on day $t-l$, and $L=3$. We estimate Equation (9) by OLS for 2,000 random stock pairs with at least nine quarters of common history of quarterly earnings (472 firms), over their longest common trading history within the sample period January 1993 to June 2004. We then compute averages (AVG) of the estimated coefficients within each quintile of firm pairs sorted according to either their absolute earnings Spearman correlations ($|\rho_{i,h}|$, EPS CORR, in Panel A), their Fama and French (1988) industry (in Panel B), or their absolute gross return Spearman correlations ($|rho_{i,h}|$, RET CORR, in Panel C) from the lowest to the highest, as well as over the fraction of the pairs (%*) for which those estimates (ESTIMATE*) are statistically significant at the 10% level (with Newey–West standard errors).

	Panel A: Sorting by CORR = EPS CORR						Panel B: Industry sort		Panel C: Sorting by CORR = RET CORR					
		Quintiles of EPS CORR								Quintiles of RET CORR				
	Total	Low	2	3	4	High	Same Industry	Different Industry	Total	Low	2	3	4	High
AVG CORR	0.245	0.040	0.124	0.214	0.330	0.518	0.279	0.239	0.214	0.126	0.172	0.205	0.240	0.328
%*	43%	0%	0%	17%	100%	100%	50%	42%	100%	100%	100%	100%	100%	100%
AVG CORR*	0.411	n.a.	n.a.	0.255	0.330	0.518	0.435	0.406	0.214	0.126	0.172	0.205	0.240	0.328
AVG DIRECT IMPACT	40.06	38.45	38.31	41.18	41.59	40.80	41.02	39.89	40.06	45.04	40.07	40.31	39.58	35.32
%*	93%	93%	93%	95%	92%	92%	89%	93%	93%	93%	93%	94%	94%	91%
AVG DIRECT IMPACT*	42.58	40.73	40.51	43.04	44.82	43.82	45.12	42.14	42.58	47.75	42.42	42.51	41.63	38.45
AVG CROSS IMPACT	6.88	6.70	6.35	7.15	7.29	6.93	7.75	6.73	6.88	7.11	6.88	7.03	6.37	7.02
%*	16%	17%	13%	18%	17%	13%	17%	15%	16%	17%	16%	17%	11%	16%
AVG CROSS IMPACT*	15.00	13.83	14.00	15.49	16.00	15.53	15.76	14.85	15.00	15.88	15.13	14.71	15.62	13.82

Earnings correlations may not fully capture the extent to which the fundamentals of two random firms are correlated (or are perceived to be correlated by market participants), for example, because of accounting conventions and practices or unobservable economic linkages. Fundamental linkages are presumably stronger among firms in the same industry. In our sample, only 15% (303) of the 2,000 random stock pairs in Panel A of Table IV is made of firms classified in the same Fama and French's (1988) industry grouping (listed in Section 4.1.a). Yet, Panel B of Table IV shows that Panel A's estimates of cumulative pairwise absolute cross-price impact in Equation (9) are similarly (economically and statistically) significant within both *same*-industry and *different*-industry random stock pairs.[11] Alternatively, stock return correlations are often used to measure fundamental similarities among firms. In our setting, conditioning the above tests to those correlations is less than ideal as both prices and order flow are jointly determined in equilibrium, while our model implies a relationship between equilibrium cross-price impact (Λ) and the exogenous covariance matrix of the terminal payoffs of the traded assets (Σ_v, see Equation (4)). With this caveat in mind, we find that sorting estimated cross-price impact among the random stock pairs in Panel A of Table IV by either the absolute Spearman correlation of their daily gross returns ($|rho_{i,h}|$, in Panel C of Table IV) or the absolute Spearman correlation of their CAPM-adjusted returns (unreported) leads to the same inference.

Overall, these results suggest that there is persistent, informational cross-price impact at the level of individual U.S. stocks over the sample period 1993–2004.

4.2 THE INFORMATIONAL ROLE OF STRATEGIC CROSS-TRADING

The evidence reported so far provides support for the main equilibrium implication of our model, that is, that the equilibrium matrix of direct and cross-asset price impact be nondiagonal in an interconnected economy as a result of the strategic cross-trading activity of sophisticated speculators. In particular, Tables III and IV indicate that cross-industry and cross-stock net order flow in the U.S. equity market have significant, persistent (i.e., informational), and often negative cross-price impact, even among less closely related industries or stocks. In this section, we test two additional predictions of our model resulting from the informational role of speculators' strategic

[11] Qualitatively similar inference also ensues from including a stock's *own* industry portfolio returns and order flow in the basic specification of Equation (9), to control for own industry-level public information and portfolio rebalancing activity.

cross-trading activity. These predictions are unique to our model, that is, cannot be generated by any of the alternative theories of trade and price co-formation discussed in Sections 1 and 3 (and assessed in Section 5). Their empirical validation would therefore provide further support for our model.

4.2.a. *Marketwide Information Heterogeneity*

The first prediction (from Corollary 1) states that, ceteris paribus, equilibrium direct and absolute cross-price impact are increasing in the marketwide heterogeneity of speculators' private information (i.e., are decreasing in ρ) because the latter makes their strategic direct and cross-trading activity more cautious and the MMs more vulnerable to adverse selection. We test this prediction parsimoniously by amending the regression models of Equations (8) and (9) to include the cross-products of direct and cross-asset order imbalance with either the average dispersion of analyst EPS forecasts (SDLTEPS_m) or the average standardized dispersion of macroeconomic forecasts (SDMMS_m), described in Section 3.2.

We estimate the following regression models,

$$\begin{aligned} r_{n,t} = {} & \alpha_n + \beta_n r_{M,t} + \sum\nolimits_{j=1}^{10} \sum\nolimits_{l=1}^{L} \gamma_{nj,l} r_{j,t-l} + \sum\nolimits_{j=1}^{10} \sum\nolimits_{l=0}^{L} \lambda_{nj,l} \omega_{j,t-l} \\ & + \sum\nolimits_{j=1}^{10} \sum\nolimits_{l=0}^{L} \lambda^{x}_{nj,l} X_t \omega_{j,t-l} + \varepsilon_{n,t}, \end{aligned} \tag{10}$$

and

$$\begin{aligned} r_{i,t} = {} & \alpha_i + \beta_i r_{M,t} + \sum\nolimits_{l=1}^{L} \gamma_{ii,l} r_{i,t-l} + \sum\nolimits_{l=1}^{L} \gamma_{ih,l} r_{h,t-l} + \sum\nolimits_{l=0}^{L} \lambda_{ii,l} \omega_{i,t-l} \\ & + \sum\nolimits_{l=0}^{L} \lambda_{ih,l} \omega_{h,t-l} + \sum\nolimits_{l=0}^{L} \lambda^{x}_{ii,l} X_t \omega_{i,t-l} \\ & + \sum\nolimits_{l=0}^{L} \lambda^{x}_{ih,l} X_t \omega_{h,t-l} + \varepsilon_{i,t}, \end{aligned} \tag{11}$$

where the variable X_t is either SDLTEPS_m or SDMMS_m. As clear from Equations (10) and (11), the scale of X_t (and the sign of $\lambda_{nj,l}$ and $\lambda_{ih,l}$) affects the scale (and sign) of the estimates for the cross-product coefficients. To ease the interpretation of the results, we compute (and report in Tables V and VI) the differences between OLS estimates of direct and absolute cross-price impact in days characterized by historically high information heterogeneity (low ρ)—that is, for X_t at the top 70th percentile of its empirical distribution ($X_{t,\,70\text{th}}$)—and those same estimates in days characterized by historically low information heterogeneity (high ρ)—that is, for X_t at the bottom 30th percentile of its empirical distribution ($X_{t,\,30\text{th}}$). Specifically, we report estimates of $\left|\sum_{l=0}^{L} \lambda_{nj,l} + \sum_{l=0}^{L} \lambda^{x}_{nj,l} X_{t,\,70\text{th}}\right|$

Table V. Marketwide information heterogeneity: direct and cross-industry price impact

This table reports the differences (ΔIMPACT, in column) between estimates of persistent direct and absolute cross-industry price impact in days characterized by historically high and low information heterogeneity for each industry portfolio (in row), from the regression model of Equation (10):

$$r_{n,t} = \alpha_n + \beta_n r_{M,t} + \sum_{j=1}^{10} \sum_{l=1}^{L} \gamma_{nj,l} r_{j,t-l} + \sum_{j=1}^{10} \sum_{l=0}^{L} \lambda_{nj,l} \omega_{j,t-l} + \sum_{j=1}^{10} \sum_{l=0}^{L} \lambda^x_{nj,l} X_t \omega_{j,t-l} + \varepsilon_{n,t},$$

where $r_{n,t-l}$ is the equal-weighted average of daily stock returns in industry portfolio n on day $t-l$, $r_{M,t}$ is the equal-weighted daily return on all NYSE and NASDAQ stocks, $\omega_{n,t-l}$ is the equal-weighted, industry-level net order flow (net scaled number of transactions) on day $t-l$, X_t is either SDLTEPS$_m$ (the equal-weighted average of stock-level standard deviations of analyst earnings forecasts, in Panel A; see Section 3.2) or SDMMS$_m$ (the simple average of the standardized dispersion of analyst forecasts of 18 macroeconomic variables, in Panel B; see Section 3.2), and $L=3$. We compute ΔIMPACT as $\left|\sum_{l=0}^{L} \lambda_{nj,l} + \sum_{l=0}^{L} \lambda^x_{nj,l} X_{t,70\text{th}}\right| - \left|\sum_{l=0}^{L} \lambda_{nj,l} + \sum_{l=0}^{L} \lambda^x_{nj,l} X_{t,30\text{th}}\right|$, where $X_{t,70\text{th}}$ and $X_{t,30\text{th}}$ are the top 70th and bottom 30th percentile of the empirical distribution of X_t. We estimate Equation (10) by OLS over the sample period January 1993 to June 2004 (2,889 observations) in Panel A, and over the subperiod January 1993 to December 2000 (2,017 observations) for Panel B, and assess the statistical significance of the estimated coefficients with Newey–West standard errors. Coefficient estimates are multiplied by 100. R^2_a is the adjusted R^2. A *, **, or *** indicates significance at the 10%, 5%, or 1% level, respectively. A ◦ indicates that neither sum is statistically significant but their difference is, nor only one sum is statistically significant and the difference is also significant but with the opposite sign.

	Durables	Nondurables	Manufacturing	Energy	HighTech	Telecom	Shops	Health	Utilities	Other	R^2_a
Panel A: ΔIMPACT for X_t = SDLTEPS$_m$											
Durables	**1.680***	0.754	1.001	−0.189	0.634	2.088**	1.777	0.446	0.246	1.901	70%
Nondurables	0.365	**1.453***	1.309	−0.077	0.844	1.216**	0.414	−0.499	−0.042	2.804***	70%
Manufacturing	−1.276**	0.354	**3.411*****	0.287	−0.489	1.845***	0.511	0.799	−0.098	0.210	85%
Energy	0.063	−1.983	2.155	**2.809*****	0.376	−0.195	0.563	1.928	0.921	2.899	62%
HighTech	0.710	−0.190	0.568	0.104	**−0.492**	1.071*◦	1.708*◦	1.264**	0.350	−2.706**	94%
Telecom	−0.548	2.009	−2.739**	−0.249	−1.980**	**3.364*****	0.996	1.701	0.301	0.332	83%
Shops	−0.501	1.446***◦	−0.248	0.052	1.839***	1.570***	**3.694*****	−0.165	0.148	1.292	86%
Health	0.905	−0.295	−0.078	0.073	3.393***	1.063	2.166**◦	**3.096*****	0.155	0.182	83%
Utilities	−0.052	2.092*	−0.852	0.320	2.604***	0.740	−1.046	0.386	**2.470*****	2.808*◦	50%
Other	0.165	0.350	0.008	0.188	1.669***	−0.117	0.007	−0.769*◦	0.066	**0.659**	91%

(continued)

Table V. (Continued)

	Durables	Nondurables	Manufacturing	Energy	HighTech	Telecom	Shops	Health	Utilities	Other	R_a^2
Panel B: ΔIMPACT for $X_t = \text{SDMMS}_m$											
Durables	**0.084**	0.044	0.473	−0.333	0.378	0.482	−0.167	0.178	0.188	0.189	61%
Nondurables	−0.018	**0.410**	−0.004	−0.373$^{**}_{\circ}$	−0.369	0.058	−0.438	−0.774$^{**}_{\circ}$	−0.037	0.712	64%
Manufacturing	−0.107	0.613	**0.046**	0.151	−0.029	−0.160	−0.783**	−0.038	0.013	0.114	79%
Energy	−0.340	−1.057	1.192*	**−0.011**	−0.957	−0.262	−0.851	−0.479	−0.338	−0.601	54%
HighTech	0.382	−0.192	−0.285	−0.081	**0.587***	0.954$^{***}_{\circ}$	−0.329	−0.257	−0.020	−0.165	93%
Telecom	0.670*	1.097*	−0.615	0.228	0.352	**0.897****	−0.710	−0.365	−0.182	0.435	81%
Shops	−0.228	−0.005	−0.458	−0.105	−0.178	−0.253	**0.681**	−0.202	0.337***	1.368***	84%
Health	0.158	−0.551	0.166	−0.497**	−0.050	0.451	0.533	**−0.583**	−0.210	−0.675	80%
Utilities	0.035	0.056	−0.039	0.314$^{*}_{\circ}$	0.407	0.298	−0.076	−0.226	**0.503*****	−0.407	49%
Other	0.148	−0.547$^{*}_{\circ}$	0.282	−0.010	0.577*	−0.177	0.365	−0.170	0.028	**0.993*****	89%

Table VI. Marketwide information heterogeneity: direct and absolute cross-price impact for random stock pairs

This table reports the differences between estimates of persistent direct and absolute cross-stock price impact in days characterized by historically *high* and *low* information heterogeneity (ΔDIRECT IMPACT and ΔCROSS IMPACT, respectively), when accompanying a 1 standard deviation shock to the corresponding order flow ($\sigma(\omega_{i,t})$ and $\sigma(\omega_{h,t})$) in basis points (i.e., multiplied by 10,000), from the regression model of Equation (11):

$$r_{i,t} = \alpha_i + \beta_i r_{M,t} + \sum_{l=1}^{L} \gamma_{ii,l} r_{i,t-l} + \sum_{l=1}^{L} \gamma_{ih,l} r_{h,t-l} + \sum_{l=0}^{L} \lambda_{ii,l} \omega_{i,t-l} + \sum_{l=0}^{L} \lambda_{ih,l} \omega_{h,t-l} + \sum_{l=0}^{L} \lambda_{ii,l}^{x} X_t \omega_{i,t-l} + \sum_{l=0}^{L} \lambda_{ih,l}^{x} X_t \omega_{h,t-l} + \varepsilon_{i,t},$$

where $r_{i,t-l}$ is the daily returns of randomly selected stock i on day $t-l$, $r_{M,t}$ is the equal-weighted daily return on all NYSE and NASDAQ stocks, $\omega_{i,t-l}$ is the daily net order flow (net scaled number of transactions) in firm i on day $t-l$, X_t is either SDLTEPS$_m$ (the equal-weighted average of stock-level standard deviation of analyst earnings forecasts, in Panel A; see Section 3.2) or SDMMS$_m$ (the simple average of the standardized dispersion of analyst forecasts of 18 macroeconomic variables, in Panel B; see Section 3.2), and $L=3$. We compute ΔDIRECT IMPACT and ΔCROSS IMPACT as $\sigma(\omega_{i,t})\left(\sum_{l=0}^{L} \lambda_{ii,l}^{x} X_{t,70\text{th}} - \sum_{l=0}^{L} \lambda_{ii,l}^{x} X_{t,30\text{th}}\right)$ and $\sigma(\omega_{h,t})\left(\left|\sum_{l=0}^{L} \lambda_{ih,l} + \sum_{l=0}^{L} \lambda_{ih,l}^{x} X_{t,70\text{th}}\right| - \left|\sum_{l=0}^{L} \lambda_{ih,l} + \sum_{l=0}^{L} \lambda_{ih,l}^{x} X_{t,30\text{th}}\right|\right)$, respectively, where $X_{t,70\text{th}}$ and $X_{t,30\text{th}}$ are the top 70th and bottom 30th percentile of the empirical distribution of X_t. We estimate Equation (11) by OLS for 2,000 random stock pairs with at least nine quarters of common history of quarterly earnings (472 firms), over their longest common trading history within the sample period January 1993 to June 2004 in Panel A and within the subperiod January 1993 to December 2000 in Panel B. We then compute averages ("AVG") of these differences within each quintile of firm pairs sorted according to their absolute earnings Spearman correlations ($|\rho_{i,h}|$, EPS CORR) from the lowest to the highest (in Panel A of Table IV), as well as over the fraction of the pairs ("%*") for which those estimates ("ESTIMATE*") are statistically significant at the 10% level (with Newey–West standard errors).

		Quintiles of EPS CORR				
	Total	Low	2	3	4	High
AVG EPS CORR	0.245	0.040	0.124	0.214	0.330	0.518
Panel A: $X_t = \text{SDLTEPS}_m$						
AVG ΔDIRECT IMPACT	11.43	10.69	10.22	11.49	12.58	12.16
%*	34%	30%	33%	37%	35%	35%

(continued)

Table VI. (Continued)

	Total	Quintiles of EPS CORR				
		Low	2	3	4	High
AVG ΔDIRECT IMPACT*	24.91	24.93	22.99	25.87	26.67	23.92
AVG ΔCROSS IMPACT	4.87	4.84	5.57	4.36	4.77	4.79
%*	17%	19%	16%	17%	18%	14%
AVG ΔCROSS IMPACT*	16.18	13.75	17.77	12.34	18.95	18.78
Panel B: $X_t = \mathrm{SDMMS}_m$						
AVG ΔDIRECT IMPACT	−0.12	−0.13	0.40	0.16	−0.89	−0.14
%*	14%	15%	15%	14%	15%	12%
AVG ΔDIRECT IMPACT*	1.53	2.53	3.48	3.17	−1.41	−0.65
AVG ΔCROSS IMPACT	−0.69	−0.43	−0.56	−1.72	−0.15	−0.62
%*	12%	12%	12%	15%	11%	10%
AVG ΔCROSS IMPACT*	−3.28	−2.44	−1.61	−5.48	−2.73	−3.46

$-\left|\sum_{l=0}^{L}\lambda_{nj,l}+\sum_{l=0}^{L}\lambda_{nj,l}^{x}X_{t,30\text{th}}\right|$ in Table V for industry portfolios, and estimates of both $\sigma(\omega_{i,t})\left(\sum_{l=0}^{L}\lambda_{ii,l}^{x}X_{t,70\text{th}}-\sum_{l=0}^{L}\lambda_{ii,l}^{x}X_{t,30\text{th}}\right)$ and $\sigma(\omega_{h,t})$ $\left(\left|\sum_{l=0}^{L}\lambda_{ih,l}^{x}+\sum_{l=0}^{L}\lambda_{ih,l}^{x}X_{t,70\text{th}}\right|-\left|\sum_{l=0}^{L}\lambda_{ih,l}^{x}+\sum_{l=0}^{L}\lambda_{ih,l}^{x}X_{t,30\text{th}}\right|\right)$ in Table VI (in basis points) for individual stocks.[12]

Consistent with Corollary 1, the estimated differences in Tables V and VI are generally positive, large, and statistically significant—especially when market-wide information heterogeneity is proxied by SDLTEPS$_m$ (Panel A), noticeably less so when by SDMMS$_m$ (Panel B, over the subperiod 1993–2000). In particular, Panel A of Table V shows that in those circumstances, the average cumulative direct (absolute cross-price) impact on an industry's daily stock returns from a 1 standard deviation shock to its own (another industry's) order flow is 21.5 (13.9) basis points greater when SDLTEPS$_m$ is high than when SDLTEPS$_m$ is low. For example, daily Manufacturing stock returns increase on average by 43.9 basis points from a 1 standard deviation shock to that industry's order flow when SDLTEPS$_m$ is high, but by just 19.7 basis points if that shock takes place while SDLTEPS$_m$ is low. Similarly, in Panel B of Table V, Telecom stock returns are generally insensitive to trading activity in Durables stocks unless when SDMMS$_m$ is high, in which case those returns decrease by 12.8 basis points in response to a 1 standard deviation shock to order imbalance in Durables stocks despite the correlation of their quarterly EPS being low and statistically insignificant (0.128 in Table II).

Table VI further indicates that direct and cross-price impact among random stock pairs are significantly higher (more often than if due to chance) in days when SDLTEPS$_m$ is larger than average, both over the entire sample and within nearly all of the earnings correlation quintiles.[13] For instance, Panel A of Table VI shows that when SDLTEPS$_m$ is high, the

[12] In these and similar subsequent tables, estimated cross-price impact coefficients occasionally change sign depending on the magnitude of X_t. In those circumstances, we report the *distance* between these estimates at $X_{t,70\text{th}}$ and $X_{t,30\text{th}}$ and sign it depending on its accordance with the model. We also mark differences (or distances) of sums of estimated price impact coefficients with the subscript "$\circ$" when (1) neither sum is statistically significant but their difference (or distance) is; or (2) only one sum is statistically significant and the difference (or distance) is also significant but with the opposite sign.

[13] Accordingly, as in Table IV, unreported t-tests strongly reject the null hypothesis that the average of these estimated differences in Panel A of Table VI is zero within all (and each quintile of) random stock pairs. As mentioned earlier, the evidence is inconclusive when $X_t = SDMMS_m$ (Panel B, over 1993–2000), perhaps suggesting that dispersion of beliefs about macroeconomic variables does not meaningfully affect strategic cross-equity trading.

daily returns of a randomly selected stock move on average by 24.9 basis points more than when $\mathrm{SDLTEPS}_m$ is low from a 1 standard deviation shock to its own order flow (in the 34% of the cases in which those differences are statistically significant) and by 16.2 basis points more from a 1 standard deviation shock to the order flow of another randomly selected stock (in the 17% of the cases in which those differences are statistically significant). Shocks to $\mathrm{SDLTEPS}_m$ have similarly significant effects on cumulative direct and cross-price impact even among random stock pairs displaying the lowest absolute earnings correlations (in the first such quintile; Panel A of Table VI). Unreported estimates of contemporaneous direct and absolute cross-price impact in Equations (10) and (11) yield even stronger inference. For example, estimated differences between direct (cross-price) impact among random stock pairs in days characterized by historically high and low dispersion of analyst earnings forecasts—that is, $\lambda_{ii,0}$ ($\lambda_{ih,0}$) of Equation (11) in high and low $X_t = \mathrm{SDLTEPS}_m$ days—average 40.9 (14.9) basis points in the 84% (22%) of the cases in which they are statistically significant.

Thus, our evidence suggests that direct and absolute cross-price impact is generally higher when the extent of information heterogeneity among speculators is high, as implied by our model.[14]

4.2.b. *Marketwide Number of Speculators*

The second prediction (also from Corollary 1) states that, ceteris paribus, the more numerous speculators are in the economy (higher M), the less cautiously they trade with their private signals, the less severe adverse selection risk becomes for uninformed market makers in all assets, hence ultimately the lower are both direct and absolute cross-price impact of net order flow. To evaluate this argument, we estimate the amended regression models of Equations (10) and (11) after allowing for the cross-product of direct and cross-asset order flow with ANA_m, the average firm-level analyst coverage (described in Section 3.2), that is, by setting $X_t = \mathrm{ANA}_m$. Again we report the differences between OLS estimates of direct and absolute cross-price impact in days characterized by historically *large* and *small* number of

[14] Further support for our model comes from sorting the estimates of cross-price impact among random stock pairs of Table IV in quintiles of their average sample-wide dispersion of analyst EPS forecasts. Ceteris paribus, cross-price impact may be larger among asset pairs with greater information heterogeneity (see also Pasquariello, 2007). Consistently, this (unreported) analysis shows not only that random stock pairs with the highest pairwise information heterogeneity display the largest cross-price impact estimates (e.g., 6.6 bps larger than in the lowest such quintile, when statistically significant), but also that these estimates are nearly monotonically increasing across pairwise information heterogeneity quintiles (despite displaying only small differences in pairwise absolute earnings correlation).

speculators—that is, for ANA_m at the top 70th and at the bottom 30th percentiles of its empirical distribution ($\mathrm{ANA}_{t,70\mathrm{th}}$ and $\mathrm{ANA}_{t,30\mathrm{th}}$)—in Panel A of Tables VII and VIII.

Consistent with Corollary 1, the estimated differences in these tables are generally negative and often statistically significant, albeit less so than in Tables V and VI. For instance, Panel A of Table VII shows that, on average, daily Shops stock returns decrease by 20.7 basis points from a 1 standard deviation shock to HighTech stocks' order flow when ANA_m is low, but by only 9.2 basis points if that shock occurs when ANA_m is high. Similarly, it is in days when ANA_m is historically low that the daily stock returns of many of the industries in our sample (especially Shops and Other) display the greatest sensitivity to cross-industry trading activity. Stronger inference ensues from unreported estimates of contemporaneous direct and absolute cross-price impact. For example, nearly 30% of estimated differences between these coefficients for industry portfolios in days characterized by historically high and low number of analysts—that is, $\lambda_{nj,0}$ of Equation (10) in high and low $X_t = \mathrm{ANA}_m$ days—are statistically significant, and more than 70% of those are negative, as predicted by Corollary 1.

However, Panel A of Table VIII indicates that average estimates of both direct and cross-price impact among random stock pairs are generally insensitive to the number of speculators in the economy, except perhaps when the absolute correlation of their earnings is low.[15] Although commonly used in the literature (e.g., Chordia, Huh, and Subrahmanyam, 2007), market-wide analyst coverage may be only an "imperfect proxy" for the number of privately informed speculators in the U.S. stock market (Brennan and Subrahmanyam, 1995, p. 362). Alternative proxies (described in Section 3.2) include the fraction of U.S. shares held by institutional investors ($\mathrm{FRACINST}_y$, which is model-free) or the intensity of informed trading (INTINF_y, based on the PIN model of Easley and O'Hara, 1992). We find that estimates of the cross-products of direct and cross-industry or cross-stock order imbalance with either $X_t = \mathrm{FRACINST}_y$ (in Panel B of Tables VII and VIII) or $X_t = \mathrm{INTINF}_y$ (unreported) in Equations (10) and (11) are broadly consistent in sign (i.e., negative) and magnitude with the evidence described above. For instance, Panel B of Tables VII and VIII indicates that estimated direct and cross-price impact among industries (random stock

[15] For example, when ANA_m is low, a 1 standard deviation shock to the order flow of a randomly selected stock within the first earnings correlation quintile of firms significantly moves that stock's daily returns (in 19% of the random pairs) by an average of almost 3 basis points less than when ANA_m is high. Accordingly, in unreported analysis, we find only estimates of *contemporaneous* price impact to uniformly decline in ANA_m.

Table VII. Marketwide number of speculators: direct and cross-industry price impact

This table reports the differences (ΔIMPACT, in column) between estimates of persistent direct and absolute cross-industry price impact in days characterized by a historically *high* and *low* number of speculators for each industry portfolio (in row), from the regression model of Equation (10):

$$r_{n,t} = \alpha_n + \beta_n r_{M,t} + \sum_{j=1}^{10} \sum_{l=1}^{L} \gamma_{nj,l} r_{j,t-l} + \sum_{j=1}^{10} \sum_{l=0}^{L} \lambda_{nj,l} \omega_{j,t-l} + \sum_{j=1}^{10} \sum_{l=0}^{L} \lambda^{x}_{nj,l} X_t \omega_{j,t-l} + \varepsilon_{n,t},$$

where $r_{n,t-l}$ is the equal-weighted average of daily stock returns in industry portfolio n on day $t-l$, $r_{M,t}$ is the equal-weighted daily return on all NYSE and NASDAQ stocks, $\omega_{n,t-l}$ is the equal-weighted, industry-level net order flow (net scaled number of transactions) on day $t-l$, X_t is either ANA_m (the equal-weighted average of analyst coverage for each stock in TAQ, in Panel A; see Section 4.2.b) or $\mathrm{FRACINST}_y$ (the equal-weighted average of the fraction of U.S. shares outstanding held by mutual funds, in Panel B; see Section 4.2.b), and $L=3$. We compute ΔIMPACT as $\left|\sum_{l=0}^{L} \lambda_{nj,l} + \sum_{l=0}^{L} \lambda^{x}_{nj,l} X_{t,70\text{th}}\right| - \left|\sum_{l=0}^{L} \lambda_{nj,l} + \sum_{l=0}^{L} \lambda^{x}_{nj,l} X_{t,30\text{th}}\right|$, where $X_{t,70\text{th}}$ and $X_{t,30\text{th}}$ are the top and bottom 30th percentile of the empirical distribution of X_t. We estimate Equation (10) by OLS over the sample period January 1993 to June 2004 (2,889 observations), and assess the statistical significance of the estimated coefficients with Newey–West standard errors. Coefficient estimates are multiplied by 100. R_a^2 is the adjusted R^2. A *, **, or *** indicates significance at the 10%, 5%, or 1% level, respectively. A ∘ indicates that neither sum is statistically significant but their difference is, nor only one sum is statistically significant and the difference is also significant but with the opposite sign.

	Durables	Nondurables	Manufacturing	Energy	HighTech	Telecom	Shops	Health	Utilities	Other	R_a^2
Panel A: ΔIMPACT for $X_t = \mathrm{ANA}_m$											
Durables	**−0.349**	−0.938	0.801	−0.224	−1.488**	−0.056	2.652**∘	−0.686	0.186	−1.973*∘	69%
Nondurables	0.337	**−0.316**	0.898	−0.050	−0.781*	0.191	0.577	0.120	−0.252	−0.434	69%
Manufacturing	−0.134	−0.088	**1.480****	0.121	−0.498	0.042	0.942	−0.332	0.295	0.480	84%
Energy	−0.293	1.159	−0.531	**−0.146**	−0.028	1.441*∘	0.327	−0.441	1.249**	−0.147	55%
HighTech	0.185	−0.121	0.281	−0.149	**−1.380*****	−0.019	−0.971	−0.406	−0.540*	−0.109	93%
Telecom	0.211	−2.039***	1.155	−0.308	0.622	**0.092**	−1.172	−0.698	−0.627	0.878	82%
Shops	−0.687*	1.172**∘	1.489***	−0.100	−1.567***	0.123	**−0.900**	0.792*∘	0.136	−0.270	85%
Health	0.762	0.558	1.559**	0.010	−0.407	0.438	1.183*∘	**−1.261****	−0.564*∘	0.003	82%
Utilities	1.638**	−1.008	1.967**	0.407	0.400	−0.503	1.644	−0.656	**−1.259*****	0.906	46%
Other	0.334	−0.789**∘	0.445	−0.212	−1.262***	0.022	0.053	0.215	−0.376**	**−0.730**	91%

(continued)

Table VII. (Continued)

	Durables	Nondurables	Manufacturing	Energy	HighTech	Telecom	Shops	Health	Utilities	Other	R_a^2
Panel B: ΔIMPACT for X_t = FRACINST$_y$											
Durables	−**0.913**	−1.049	0.240	0.409	0.603	0.839	$-2.331^{*}_{\circ}$	0.119	−0.744	−0.615	69%
Nondurables	−0.055	−**0.912**	−0.829	0.120	1.410**	−0.650	−1.090	−1.861***	0.023	−0.984	67%
Manufacturing	0.704	−0.780	−**2.540*****	−0.356	0.428	−1.855***	−0.459	−0.173	0.335	0.242	84%
Energy	−2.407**	1.482	2.410*	−**0.836**	0.884	−0.948	−1.265	−0.359	−0.641	−0.554	56%
HighTech	−0.944	−0.428	0.195	0.214	**1.242****	−0.149	−1.134	0.889*	−0.945***	0.638	93%
Telecom	−0.885	−2.370***	0.934	0.252	0.984	−**4.859*****	−0.480	$1.643^{**}_{\circ}$	−0.482	−2.305*	82%
Shops	1.146**	$-1.014^{*}_{\circ}$	−0.746	−0.033	−0.589	−1.581***	−**2.444*****	0.023	0.155	−1.065	85%
Health	−0.928	−1.158	−1.382	−1.062***	−2.451***	−1.357*	−0.334	−**0.517**	$-0.778^{*}_{\circ}$	0.109	82%
Utilities	0.389	−1.767*	1.480	−0.062	−4.025***	-0.343	0.918	−0.677	−**2.713*****	−1.704	46%
Other	0.580	−0.796	0.207	−0.199	−0.914*	−0.502	0.709	−0.128	$-0.594^{**}_{\circ}$	−**0.779**	90%

Table VIII. Marketwide number of speculators: direct and absolute cross-price impact for random stock pairs

This table reports the differences between estimates of persistent direct and absolute cross-stock price impact in days characterized by a historically *high* and *low* number of speculators, (ΔDIRECT IMPACT and ΔCROSS IMPACT, respectively), when accompanying a 1 standard deviation shock to the corresponding order flow ($\sigma(\omega_{i,t})$ and $\sigma(\omega_{h,t})$) in basis points (i.e., multiplied by 10,000), from the regression model of Equation (11):

$$r_{i,t} = \alpha_i + \beta_i r_{M,t} + \sum_{l=1}^{L} \gamma_{ii,l} r_{i,t-l} + \sum_{l=1}^{L} \gamma_{ih,l} r_{h,t-l} + \sum_{l=0}^{L} \lambda_{ii,l}\omega_{i,t-l} + \sum_{l=0}^{L} \lambda_{ih,l}\omega_{h,t-l} + \sum_{l=0}^{L} \lambda_{ii,l}^{x} X_m \omega_{i,t-l} + \sum_{l=0}^{L} \lambda_{ih,l}^{x} X_m \omega_{h,t-l} + \varepsilon_{i,t},$$

where $r_{i,t-l}$ is the equal-weighted average of the daily returns of randomly selected stock i on day t–l, $r_{M,t}$ is the equal-weighted daily return on all NYSE and NASDAQ stocks, $\omega_{i,t-l}$ is the daily net order flow (net scaled number of transactions) in firm i on day t–l, X_t is either ANA_m (the equal-weighted average of analyst coverage for each stock in TAQ, in Panel A; see Section 4.2.b) or $\mathrm{FRACINST}_y$ (the equal-weighted average of the fraction of U.S. shares outstanding held by mutual funds, in Panel B; see Section 4.2.b), and $L=3$. We compute ΔDIRECT IMPACT and ΔCROSS IMPACT as $\sigma(\omega_{i,t})\left(\sum_{l=0}^{L} \lambda_{ii,l}^{x} X_{t,70\text{th}} - \sum_{l=0}^{L} \lambda_{ii,l}^{x} X_{t,30\text{th}}\right)$ and $\sigma(\omega_{h,t})\left(\left|\sum_{l=0}^{L} \lambda_{ih,l} + \sum_{l=0}^{L} \lambda_{ih,l}^{x} X_{t,70\text{th}}\right| - \left|\sum_{l=0}^{L} \lambda_{ih,l} + \sum_{l=0}^{L} \lambda_{ih,l}^{x} X_{t,30\text{th}}\right|\right)$, respectively, where $X_{t,70\text{th}}$ and $X_{t,30\text{th}}$ are the top 70th and bottom 30th percentile of the empirical distribution of X_t. We estimate Equation (11) by OLS for 2000 random stock pairs with at least nine quarters of common history of quarterly earnings (472 firms), over their longest common trading history within the sample period January 1993 to June 2004. We then compute averages ("AVG") of estimates within each quintile of firm pairs sorted according to their absolute earnings Spearman correlations ($|\rho_{i,h}|$, EPS CORR) from the lowest to the highest, as well as over the fraction of the pairs ("%*") for which those estimates ("ESTIMATE*") are statistically significant at the 10% level (with Newey–West standard errors).

		Quintiles of EPS CORR				
	Total	Low	2	3	4	High
AVG EPS CORR	0.245	0.040	0.124	0.214	0.330	0.518
Panel A: $X_t = \mathrm{ANA}_m$						
AVG ΔDIRECT IMPACT	−0.13	−0.50	−0.27	−0.31	0.16	0.28
%*	18%	19%	18%	19%	20%	16%
AVG ΔDIRECT IMPACT*	−0.41	−3.13	0.42	0.64	−0.37	0.49
AVG ΔCROSS IMPACT	0.85	0.58	1.51	0.64	0.86	0.67
%*	14%	15%	14%	12%	14%	15%
AVG ΔCROSS IMPACT*	4.01	1.43	5.50	3.56	2.94	6.45
Panel B: $X_t = \mathrm{FRACINST}_y$						
AVG ΔDIRECT IMPACT	−4.30	−4.23	−4.41	−3.92	−3.75	−5.17
%*	23%	22%	25%	24%	21%	24%
AVG ΔDIRECT IMPACT*	−8.86	−8.55	−9.29	−9.14	−4.63	−12.12
AVG ΔCROSS IMPACT	−1.90	−2.15	−3.53	−0.88	−1.61	−1.34
%*	16%	17%	19%	17%	15%	13%
AVG ΔCROSS IMPACT*	−5.31	−3.17	−9.56	−3.64	−6.88	−2.45

pairs) often decline in the fraction of institutional ownership of U.S. stocks—for example, by 25 (9) and 13 (5) basis points per 1 standard deviation shock to direct and cross-industry (cross-stock) order flow, respectively, when statistically significant.[16] Thus, Tables VII and VIII suggest that the magnitude of direct and absolute cross-industry price impact is lower in days when the market-wide number of speculators is large, as implied by our model.

Overall, the earlier results provide additional support for our model, for they indicate that direct and cross-price impact in the U.S. equity market may be related to the informational role of the strategic direct and cross-trading activity of better informed speculators in that market.

5. Robustness Analysis

In this section, we gauge the robustness of our inference to several alternative specifications of our empirical strategy. Our analysis indicates that this inference is indeed robust.

5.1 ALTERNATIVE THEORIES OF CROSS-TRADING

We begin by assessing the importance of alternative theories of the relationship between cross-trading and cross-price impact for the evidence presented in Section 4. As in our model, these theories also emphasize the role of financial linkages (rather than real ones) in stock price co-formation. Yet, they propose alternative mechanisms—for example, related to the extent and dynamics of market-wide fundamental uncertainty, risk aversion, and financial constraints—potentially leading to equilibrium cross-price impact. Tables III to VIII do not explicitly control for any of these mechanisms. We do so in this section.

Current literature groups these alternative channels of transmission into several, often related categories (see the discussion in Pasquariello, 2007). The

[16] In addition, unreported t tests reject the null hypothesis that the average estimated differences in direct and cross-price impact for $X_t = \mathrm{FRACINST}_y$ (in Panel B of Table VIII) are zero within all (and most quintiles of) random stock pairs; the *contemporaneous* estimates are also uninformly negative, yet even larger and more often statistically significant. In unreported analysis, positive shocks to the fraction of informed ownership of U.S. stocks have similarly nontrivial negative effects on cumulative direct and cross-industry price impact, for example averaging 11 and 12 basis points per 1 standard deviation shock to direct and cross-industry order flow, respectively (when statistically significant in Equation (10)). Average estimates of cross-product coefficients of random stock-level order imbalance with $X_t = \mathrm{INTINF}_y$ in Equation (11) are less often significant.

first one, the correlated information channel (King and Wadhwani, 1990; Chan, 1993), is based on the idea that in the presence of information asymmetry among investors, cross-trading activity motivated by correlated information shocks may lead to cross-price impact. By construction, this mechanism precludes cross-price impact among fundamentally unrelated assets because it crucially depends on uninformed investors' cross-inference about traded assets' related terminal payoffs (in absence of financial intermediation). Both our model and empirical results suggest otherwise. However, this mechanism also implies that greater market-wide information asymmetry may lead to higher direct and absolute cross-price impact. We measure market-wide information asymmetry about U.S. stocks' future prospects with two alternative proxies. The first one is EPSVOL_q (Figure 2f), the equal-weighted average of firm-level EPS volatility in calendar quarter q (over the most recent eight quarters), as in Chordia, Huh, and Subrahmanyam (2007). The second one is EURVOL_m (Figure 2g), the monthly average (to smooth daily variability) of daily Eurodollar implied volatility from Bloomberg. EURVOL_m is a commonly used measure of market participants' perceived uncertainty surrounding U.S. monetary policy, an important source of fundamental uncertainty in the U.S. stock market (Bernanke and Kuttner, 2005).

Within the second category of theories, Kodres and Pritsker (2002) argue that in economies populated by uninformed investors learning from prices, the portfolio rebalancing activity of privately informed, price-taking, investors—driven by risk aversion—may induce contemporaneous price covariance and cross-price impact, even among assets with uncorrelated payoffs. As mentioned in Section 4.1.a, both this intuition and the potential trading patterns ensuing from style investing (Barberis, Shleifer, and Wurgler, 2005) or correlated informational trading motivate the inclusion of contemporaneous market returns $r_{M,t}$ in the basic empirical specifications of Equations (8) and (9). We further measure the extent and dynamics of market-wide risk aversion, or risk appetite in the U.S. stock market with a model-free proxy suggested by Bollerslev, Tauchen, and Zhou (2009), RISKAV_m. This proxy (Figure 2h) is computed as the monthly difference between the end-of-month Chicago Board Options Exchange (CBOE)'s VIX index of implied volatility of S&P500 options with a fixed 30-day maturity (VIX_m) and the realized volatility of 5-minute S&P500 returns ($r_{SP,\tau}$) within the month ($\mathrm{RV}_m = \sqrt{\sum_{\tau \in m} r^2_{\mathrm{SP},\tau}}$).

A third set of studies models the cross-trading activity of speculators, even across fundamentally unrelated assets, as the equilibrium outcome of correlated liquidity shocks due to financial constraints (Kyle and Xiong, 2001). In the presence of those shocks, speculators' trading activity may also lead to equilibrium cross-price impact by influencing the inferences and trades of other speculators and uninformed investors via prices (Bernhardt and Taub, 2008), rather than via order flow (as in our model). The most direct implication of these arguments for our analysis is that, on average, absolute cross-industry price impact may be asymmetric—that is, higher during times when borrowing, short-selling, and wealth constraints are particularly binding (e.g., during periods of liquidity crises)—and/or sensitive to the dynamics of interest rates (Shiller, 1989). We proxy for the former with a dummy, d_t^{CR}, equal to one if day t falls within any of the liquidity crisis periods listed by Chordia, Sarkar, and Subrahmanyam (2005), and equal to zero otherwise.[17] We capture the latter with a measure of time-varying risk-free interest rates, $r_{RF,t}$, the daily time series of 1-month Treasury Bill rates (from CRSP).

We assess the relevance of these arguments for our inference parsimoniously by including the cross-products of direct and cross-asset order flow with the various proxies described above in the regression models of Equations (10) and (11). We first estimate Equations (10) and (11) after replacing the information variable X_t with $\mathrm{EPSVOL}_q, \mathrm{EURVOL}_t$, RISKAV_t, $r_{RF,t}$, or d_t^{CR}, separately, that is, while ignoring both the number of speculators and their information heterogeneity. The results of this analysis, not reported here for economy of space, provide (at best) weak support for the notion that direct and absolute cross-price impact may be increasing in market-wide risk aversion, and little or no evidence of direct and cross-price impact being sensitive to fluctuations in market-wide fundamental uncertainty, cost of borrowing, or liquidity crises.

Changes in fundamental uncertainty (or information asymmetry) may also cloud the interpretation of our comparative statics tests by affecting both the market's information environment and dealers' inventory management. To evaluate this possibility, we amend Equations (10) and (11) to include the cross-products of direct and cross-asset order flow with our proxies for either information heterogeneity or the number of speculators (X_t), and either of the proxies for market-wide information asymmetry described earlier (X_t^v) as follows:

$$\begin{aligned} r_{n,t} = \alpha_n + \beta_n r_{M,t} + \sum\nolimits_{j=1}^{10} \sum\nolimits_{l=1}^{L} \gamma_{nj,l} r_{j,t-l} + \sum\nolimits_{j=1}^{10} \sum\nolimits_{l=0}^{L} \lambda_{nj,l} \omega_{j,t-l} \\ + \sum\nolimits_{j=1}^{10} \sum\nolimits_{l=0}^{L} \lambda_{nj,l}^{x} X_t \omega_{j,t-l} + \sum\nolimits_{j=1}^{10} \sum\nolimits_{l=0}^{L} \lambda_{nj,l}^{v} X_t^v \omega_{j,t-l} + \varepsilon_{n,t}, \end{aligned} \tag{12}$$

[17] These periods are: March 1 1994 to May 31 1994 (U.S. bond market crisis); July 2 1997 to December 31 1997 (Asian crisis); and July 6 1998 to December 31 1998 (Russian default crisis).

for the 10 industries listed in Section 4.1.a, and

$$\begin{aligned} r_{i,t} =& \alpha_i + \beta_i r_{M,t} + \sum\nolimits_{l=1}^{L} \gamma_{ii,l} r_{i,t-l} + \sum\nolimits_{l=1}^{L} \gamma_{ih,l} r_{h,t-l} + \sum\nolimits_{l=0}^{L} \lambda_{ii,l} \omega_{i,t-l} \\ &+ \sum\nolimits_{l=0}^{L} \lambda_{ih,l} \omega_{h,t-l} + \sum\nolimits_{l=0}^{L} \lambda_{ii,l}^{x} X_t \omega_{i,t-l} + \sum\nolimits_{l=0}^{L} \lambda_{ih,l}^{x} X_t \omega_{h,t-l} \\ &+ \sum\nolimits_{l=0}^{L} \lambda_{ii,l}^{v} X_t^{v} \omega_{i,t-l} + \sum\nolimits_{l=0}^{L} \lambda_{ih,l}^{v} X_t^{v} \omega_{h,t-l} + \varepsilon_{i,t}, \end{aligned} \tag{13}$$

for random stock pairs, where X_t is $\mathrm{SDLTEPS}_m$, SDMMS_m, ANA_m, or $\mathrm{FRACINST}_y$ while X_t^v is either EPSVOL_q or EURVOL_t. As in Section 4.2, we compute (but again do not report here) the resulting differences between absolute OLS estimates of direct and cross-price impact from Equations (12) and (13) when the corresponding information variable X_t is historically high ($X_{t,70\mathrm{th}}$) and when X_t is historically low ($X_{t,30\mathrm{th}}$). We find these differences to be broadly consistent in sign, magnitude, and (economic and statistical) significance with those in Tables V to VIII among both industry portfolios and random stock pairs, that is, to lead to qualitatively similar inference. Hence, we conclude that the evidence presented in Section 4 in support of our model is robust to allowing for direct and cross-price impact to respond to fluctuations in fundamental volatility, risk aversion, or financial constraints.

5.2 PRICE VERSUS ORDER FLOW OBSERVABILITY

We also consider the empirical relevance of price observability for equilibrium cross-price impact from strategic multi-asset trading, as suggested by the theoretical work of Bernhardt and Taub (2008). Albeit requiring exogenous correlated noise trading (absent from our setting), this channel is potentially complementary to the transmission mechanism described in our model, based instead on order flow observability. For this analysis, we exploit an important feature of the NYSE, namely the fact that NYSE dealers (specialists) specialize in nonintersecting subsets of the traded stocks (Corwin, 1999; Hasbrouck, 2007). To the extent that cross-order flow observability is likely to be the highest for stocks dealt by the same specialist, ceteris paribus we would expect average cross-price impact to be higher among those stocks than among stocks dealt by different specialists.

To test for this argument, we use specialist information on NYSE-listed stocks from the NYSE Post and Panel File (Coughenour and Deli, 2002). This information—available to us between November 2001 and June 2004—is accessible to all market participants and allows us to identify the specialists dealing multiple NYSE stocks by matching those stocks' Post and

Panel locations on the NYSE trading floor. We then estimate and compare cumulative direct and pairwise absolute cross-price impact (Equation (9)) accompanying a 1 standard deviation shock to direct and cross-stock order imbalance for two sets of stock pairs over the subperiod November 2001 to June 2004: (1) all stock pairs always dealt by the same specialist and specialist firm during that interval (80 pairs); and (2) the same number of random stock pairs always dealt by a different specialist and specialist firm during that interval.[18] We report averages of these estimates for each earnings correlation quintile of those stocks in Panels A (random NYSE-only stock pairs dealt by the same specialist) and B (random NYSE-only stock pairs dealt by different specialists) of Table IX. Importantly for this comparison, the random stocks in these two panels do not otherwise significantly differ across a variety of firm-level characteristics.[19]

Table IX highlights an important result. Average estimated direct and especially absolute cross-price impact among stock pairs dealt by the same specialist (i.e., with the highest cross-order flow observability) are as often statistically significant but greater in magnitude than the corresponding estimates among stock pairs dealt by different specialists (i.e., with more limited cross-order flow observability)—as implied by our model—over the entire sample and within most earnings correlation quintiles. These differences are also economically significant. For example, Table IX shows that a 1 standard deviation shock to a stock's order imbalance moves its own (another stock's) daily returns by an average of 6.3 (1.7) basis points more—that is, by 21% (7%) more—if both stocks are dealt by the same specialist than if they are dealt by different specialists.[20] These estimated

[18] Consistent with Section 4.1.b, we restrict our analysis to NYSE-only stock pairs with at least nine quarters of common history of quarterly earnings over our full sample period. These requirements yield 371 firms, with similar size and industry distributions to those of the 472 NYSE and NASDAQ stocks in the firm-level analysis of Table IV. We then estimate direct and cross-price impact over these pairs' longest common trading history in TAQ between November 2001 and June 2004.

[19] For instance, average absolute pairwise EPS correlations are very similar in aggregate and across quintiles of random pairs of same-specialist (Panel A) and different-specialist NYSE stocks (Panel B) in Table IX. Further unreported analysis indicates that, as in Coughenour and Saad (2004) and Chakrabarty and Moulton (2012), NYSE specialists' stock portfolios in our sample are not concentrated across industry or market capitalization. Consistently, Corwin (2004) observes that NYSE stocks are allocated among specialist firms primarily according to those firms' relative position in the *queue*—that is, the time since each received a prior allocation—rather than according to any particular stock characteristic.

[20] Inventory and shared-capital considerations may also magnify cross-price impact among stocks dealt by the same specialist (e.g., Coughenour and Saad, 2004). However, according

Table IX. Price observability: direct and absolute cross-price impact for random stock pairs

This table reports estimates of average persistent direct and cross-stock price impact (DIRECT IMPACT as $\sum_{l=0}^{L}\lambda_{ii,l}$ and $\sum_{l=0}^{L}\lambda_{hh,l}$, and CROSS IMPACT as $\left|\sum_{l=0}^{L}\lambda_{ih,l}\right|$ and $\left|\sum_{l=0}^{L}\lambda_{hi,l}\right|$) accompanying a 1 standard deviation shock to the corresponding order flow ($\sigma(\omega_{i,t})$ and $\sigma(\omega_{h,t})$) in basis points (i.e., multiplied by 10,000), from the regression model of Equation (9):

$$r_{i,t} = \alpha_i + \beta_i r_{M,t} + \sum_{l=1}^{L}\gamma_{ii,l}r_{i,t-l} + \sum_{l=1}^{L}\gamma_{ih,l}r_{h,t-l} + \sum_{l=0}^{L}\lambda_{ii,l}\omega_{i,t-l} + \sum_{l=0}^{L}\lambda_{ih,l}\omega_{h,t-l} + \varepsilon_{i,t},$$

where $r_{i,t-l}$ is the equal-weighted average of the daily returns of randomly selected stock i on day $t-l$, $r_{M,t}$ is the equal-weighted daily return on all NYSE and NASDAQ stocks, $\omega_{i,t-l}$ is the daily net order flow (net scaled number of transactions) in firm i on day $t-l$, and $L=3$. We estimate Equation (9) by OLS for 80 random pairs of NYSE stocks with at least nine quarters of common history of quarterly earnings (371 firms), over their longest common trading history within the sample period November 2001 to June 2004, as well as dealt by either the same (Panel A) or different specialist and specialist firm (Panel B). We then compute averages ("AVG") of the estimated coefficients within each quintile of firm pairs sorted according to their absolute earnings Spearman correlations ($|\rho_{i,h}|$, EPS CORR) from the lowest to the highest, as well as over the fraction of the pairs ("%*") for which those estimates ("ESTIMATE*") are statistically significant at the 10% level (with Newey–West standard errors).

	Panel A: Same specialist						Panel B: Different specialist					
		Quintiles of EPS CORR						Quintiles of EPS CORR				
	Total	Low	2	3	4	High	Total	Low	2	3	4	High
AVG EPS CORR	0.236	0.029	0.112	0.193	0.318	0.526	0.256	0.035	0.106	0.229	0.382	0.529
%*	35%	0%	0%	13%	63%	100%	53%	0%	0%	63%	100%	100%
AVG EPS CORR*	0.425	n.a.	n.a.	0.219	0.306	0.526	0.409	n.a.	n.a.	0.260	0.382	0.529
AVG DIRECT IMPACT	24.10	17.56	21.70	18.48	37.93	24.84	24.32	24.11	29.31	22.30	25.36	20.51
%*	53%	44%	50%	50%	69%	50%	71%	88%	75%	75%	75%	44%
AVG DIRECT IMPACT*	36.48	28.30	35.12	26.64	49.20	37.37	30.18	26.01	36.06	27.06	33.01	28.97
AVG CROSS IMPACT	12.98	15.38	13.77	9.94	11.16	14.56	10.82	12.14	11.52	10.42	9.51	10.51
%*	15%	25%	19%	19%	0%	13%	15%	19v	25%	19%	0%	13%
AVG CROSS IMPACT*	24.76	29.07	26.34	16.05	n.a.	26.82	23.08	27.04	24.39	20.69	n.a.	18.09

differences are as large as 16.2 (8.7) basis points—that is, as large as 49% (48%)—within the fourth (fifth) quintile of stock pairs (whose quarterly EPS correlations are most often statistically significant).[21] This evidence suggests that, consistent with our model, order flow observability may have a first-order effect on direct and cross-price impact in the U.S. stock market.

5.3 THE EVOLUTION OF U.S. STOCK MARKET STRUCTURE

Both the structure of the U.S. equity market and its information and regulatory environment have evolved considerably in the last two decades. In particular, the introduction of new trading rules and mechanisms—most notably, decimalization and the electronic limit order book—and the emergence of alternative trading platforms (known as electronic communication networks, ECNs) and of the more opaque "dark pools" have significantly affected the process of U.S. stock price formation (Bessembinder, 2003b; Hendershott and Jones, 2005).

It is plausible that these innovations may also affect the economic significance of the implications of strategic cross-trading on direct and cross-price impact, as predicated by our model. For instance, the fragmentation of trading in NYSE stocks across venues may weaken order flow observability (and its implications for price co-formation and market liquidity described in Sections 2 and 4). The adoption of rules requiring greater disclosure of private material information to the public (e.g., Regulation FD) may attenuate the informativeness of the order flow. Yet, exchange consolidation (e.g., between the NASDAQ and the INET and Brut ECNs), the increasingly widespread adoption of the electronic limit order book (e.g., within the NASDAQ's SuperMontage), and regulation aimed at increasing

to the literature (e.g., Hasbrouck, 1991), estimates of *cumulative* cross-price impact coefficients $\sum_{l=0}^{L} \lambda_{ih,l}$ in Equation (9) (like those reported in Table IX) are likely to be robust to those transient, noninformational effects. In addition, unreported estimates of *contemporaneous* cross-price impact $\lambda_{ih,0}$ in Equation (9), while more likely to be affected by inventory considerations, do *not* yield larger differentials between stocks dealt by the same or different specialists than those in Table IX.

[21] Order flow observability is also likely to be higher for NYSE stocks dealt by specialists employed by the same specialist firms (but lower than for stocks dealt by the same specialist); however, a specialist firm's efforts at managing its aggregate stock inventory are likely greater than the efforts of any individual specialist it employs (e.g., Coughenour and Saad, 2004). Alleviating the same-specialist constraint also more than doubles the number of stock pairs in the analysis, relative to Table IX. Nevertheless, grouping stocks according to their specialist firms yields qualitatively and quantitatively similar inference.

transparency within the novel trading platforms (e.g., Regulation NMS) may instead enhance order flow observability and facilitate strategic cross-trading.

To assess the relevance of these arguments for our inference, we exploit two important features of our sample. First, our sample includes stocks traded either on a hybrid, specialist-based platform—the NYSE (open outcry, dealer, and electronic limit order book)—or on a (primarily) dealer-based platform—the NASDAQ (electronic limit order book for dealers). The latter resembles more closely the platforms adopted by new ECNs (Hasbrouck, 2007). Second, our sample (January 1993 to June 2004) spans the more recent period (e.g., since 2000) when many of the aforementioned changes to U.S. equity trading began to occur (see also Stoll, 2001; Bennett and Wei, 2006; O'Hara and Ye, 2011). These features allow us to test for whether the economic significance of our inference is stable across trading platforms and over time. We estimate (and report in Table X) average cumulative direct and absolute cross-price impact—from Equation (9), in basis points per 1 standard deviation shocks to order flow (as in Table IV)—among 2,000 random pairs (1) of either NYSE-only (Panel A) or NASDAQ-only stocks (Panel B), according to their listing exchange; as well as (2) over either of two consecutive nonoverlapping subsample periods, January 1993 to December 1999 (Panel C) and January 2000 to June 2004 (Panel D).

Table X indicates that cross-price impact is salient within both the NYSE and the NASDAQ—that is, within alternative trading venues, technologies, and execution procedures—as well as throughout our sample period. For instance, Panels A and B show that cross-price impact is relatively larger among NYSE-only random stock pairs—for example, between 38% and 44% of the corresponding direct price impact, when statistically significant. Yet, cross-price impact among NASDAQ-only random stock pairs is also economically, and as often statistically, significant—for example, in that case averaging 23 basis points versus a mean direct price impact of 76 basis points. Panels C and D further suggest that the importance of cross-price impact did not diminish (and possibly increased) over our sample period—for example, averaging 19 basis points in the 1990s and 26 basis points in the first half of the 2000s (when statistically significant)—even though mean direct price impact declined. Similar inference ensues from unreported tests of the unique implications of our model related to market-wide information heterogeneity and the number of speculators (as in Section 4.2 and Tables V–VIII) for each of the above sample partitions.

Based on this evidence, we conjecture that the implications of strategic cross-trading and order flow observability for U.S. stock price

Table X. Further analysis of direct and absolute cross-price impact for random stock pairs

This table reports estimates of average persistent direct and cross-stock price impact (DIRECT IMPACT as and CROSS IMPACT as $\left|\sum_{l=0}^{L} \lambda_{ih,l}\right|$) accompanying a 1 standard deviation shock to the corresponding order flow (and $\sigma(\omega_{h,t})$) in basis points (i.e., multiplied by 10,000), from the regression model of Equation (9):

$$r_{i,t} = \alpha_i + \beta_i r_{M,t} + \sum\nolimits_{l=1}^{L} \gamma_{ii,l} r_{i,t-l} + \sum\nolimits_{l=1}^{L} \gamma_{ih,l} r_{h,t-l} + \sum\nolimits_{l=0}^{L} \lambda_{ii,l} \omega_{i,t-l} + \sum\nolimits_{l=0}^{L} \lambda_{ih,l} \omega_{h,t-l} + \varepsilon_{i,t},$$

where $r_{i,t-l}$ is the equal-weighted average of the daily returns of randomly selected stock i on day $t-l$, $r_{M,t}$ is the equal-weighted daily return on all NYSE and NASDAQ stocks, $\omega_{i,t-l}$ is the daily net order flow (net scaled number of transactions) in firm i on day $t-l$, and $L=3$. We estimate Equation (9) by OLS for 2,000 random pairs of either NYSE-only stocks (371 firms, Panel A) or NASDAQ-only stocks (103 firms, Panel B), according to their CRSP listing exchange code, with at least nine quarters of common history of quarterly earnings, over their longest common trading history within the sample period January 1993 to June 2004, as well as for 2,000 random pairs of similarly selected NYSE and NASDAQ stocks (472 firms) over either the subsample period January 1993 to December 1999 (Panel C) or the subsample period January 2000 to June 2004 (Panel D). We then compute averages ("AVG") of the estimated coefficients within each quintile of firm pairs sorted according to their absolute earnings Spearman correlations over the sample period January 1993 to June 2004 ($|\rho_{i,h}|$, EPS CORR) from the lowest to the highest, as well as over the fraction of the pairs ("%*") for which those estimates ("ESTIMATE*") are statistically significant at the 10% level (with Newey–West standard errors).

		Quintiles of EPS CORR						Quintiles of EPS CORR				
	Total	Low	2	3	4	High	Total	Low	2	3	4	High
	Panel A: NYSE-only random stock pairs						Panel B: NASDAQ-only random stock pairs					
AVG EPS CORR	0.242	0.038	0.119	0.212	0.320	0.521	0.266	0.048	0.138	0.239	0.355	0.549
%*	43%	0%	0%	15%	100%	100%	48%	0%	0%	42%	100%	100%
AVG EPS CORR*	0.409	n.a.	n.a.	0.253	0.320	0.521	0.420	n.a.	n.a.	0.270	0.355	0.549
AVG DIRECT IMPACT	31.66	31.21	31.69	31.40	33.21	30.76	75.85	78.45	78.00	75.91	77.95	68.94
%*	93%	93%	94%	94%	95%	90%	95%	99%	98%	95%	95%	89%
AVG DIRECT IMPACT*	33.49	32.95	33.35	33.00	34.71	33.42	79.86	79.21	79.86	80.26	81.75	78.09
AVG CROSS IMPACT	6.10	5.80	5.99	6.34	6.05	6.31	10.25	10.32	10.04	10.62	9.90	10.39
%*	16%	14%	19%	17%	15%	14%	18%	20%	18%	19%	17%	17%
AVG CROSS IMPACT*	13.77	13.13	12.76	14.04	14.37	14.70	22.59	22.46	22.41	23.07	22.35	22.67

(*continued*)

Table X. Continued

	Panel C: January 1993 to December 1999						Panel D: January 2000 to June 2004					
		Quintiles of EPS CORR						Quintiles of EPS CORR				
	Total	Low	2	3	4	High	Total	Low	2	3	4	High
AVG EPS CORR	0.245	0.040	0.124	0.214	0.330	0.518	0.245	0.040	0.124	0.214	0.330	0.518
%*	43%	0%	0%	17%	100%	100%	43%	0%	0%	17%	100%	100%
AVG EPS CORR*	0.411	n.a.	n.a.	0.255	0.330	0.518	0.411	n.a.	n.a.	0.255	0.330	0.518
DIRECT IMPACT	43.31	41.53	41.43	44.88	45.17	43.55	39.78	39.08	38.41	40.20	41.25	39.98
%*	90%	92%	90%	91%	88%	88%	83%	82%	84%	81%	82%	86%
AVG DIRECT IMPACT*	47.24	44.65	45.19	48.24	50.10	48.13	45.82	45.47	43.26	47.02	48.15	45.29
AVG CROSS IMPACT	8.51	8.45	7.65	8.76	9.18	8.52	11.98	11.81	12.16	11.32	12.41	12.19
%*	15%	16%	12%	16%	14%	14%	16%	17%	15%	14%	16%	15%
AVG CROSS IMPACT*	18.55	18.18	17.62	18.98	20.50	17.45	25.58	25.73	26.09	25.61	25.64	24.83

co-formation—as modeled and assessed in our prior analysis—are likely to remain significant within the rapidly evolving landscape of the U.S. equity market.

6. Related Literature

Our work is related to recent studies examining cross-stock linkages. Hasbrouck and Seppi (2001) and Hartford and Kaul (2005) find evidence of common effects in returns and order flow among Dow 30 and S&P500 stocks, respectively, and attribute most of the observed return commonality to order flow commonality. Greenwood (2005) employs a limits-to-arbitrage model and event returns around a unique redefinition of the Nikkei 225 index in Japan in April 2000 to argue that in the short run, the hedging needs of risk averse arbitrageurs may make a stock's returns sensitive to uninformed demand shocks to other stocks with correlated fundamentals. Consistently, Andrade, Chang, and Seasholes (2008) demonstrate that in a multi-asset extension of Grossman and Miller (1988), the hedging needs of risk averse liquidity providers may lead to cross-price impact of nonnformational, inelastic trading if asset payoffs are correlated, despite the absence of cross-trading. Using data from margin accounts set up by individual investors with local brokerage firms in the Taiwan Stock Exchange (TSE), Andrade, Chang, and Seasholes (2008) find support for this implication by showing that individual weekly stock returns are more positively related to trading imbalances in more related industry portfolios.

Cross-stock linkages have also been attributed to information. Motivated by a model in which an oligopolistic product market makes firm-specific news relevant to the value of all firms in that market and multi-asset trading by firm insiders is ruled out by assumption, Tookes (2008) documents that the intraday stock returns of earnings-announcing U.S. firms are sensitive to both intraday order flows and stock returns of other nonannouncing firms within the same industry. Watanabe (2008) shows that allowing for endogenous information acquisition and common shocks to GARCH-type volatility of fundamentals in the model of Caballé and Krishnan (1994) may explain why estimates of intraday direct price impact for each stock in the Dow Jones Industrial Average (DJIA) index are sensitive to lagged squared information shocks of both itself and an average of the other stocks in that index.

Our analysis differs from these studies in that we investigate, both theoretically and empirically (using transaction-level data), the properties of persistent cross-price impact in the entire U.S. stock market in the presence of strategic, informational cross-trading, even among different industries and random stock pairs.

7. Conclusions

This study presents a novel investigation of the informational role of strategic trading for the process of price co-formation in the U.S. equity market.

To motivate our empirical analysis, we develop a stylized model of multi-asset trading in the presence of strategic speculators endowed with diverse private information about the traded assets. This model, based on Kyle (1985), allows us to precisely and parsimoniously characterize the equilibrium properties of both direct and cross-price impact—the impact of trading in one asset on both its own price and the prices of other assets—when alternative theoretical channels of trade and price co-formation (inventory management, correlated information, portfolio rebalancing, correlated liquidity, and price observability) are absent by assumption. We show that, even in those circumstances, cross-price impact can be the equilibrium outcome of speculators' strategic trading activity across many assets to mask their information advantage about some other assets.

We find support for these cross-asset informational effects in a comprehensive sample of the trading activity in NYSE and NASDAQ stocks between 1993 and 2004. In particular, we report robust evidence that order flow in one stock or industry has a significant and persistent (i.e., informational) impact on daily returns of other (even fundamentally unrelated) stocks or industries. Our empirical analysis also suggests that, as implied by our model, cross-price impact is often negative and both direct and absolute cross-price impact are smaller when speculators are more numerous, greater when their information heterogeneity is higher, and greater among stocks dealt by the same specialist.

Overall, these findings indicate that cross-price impact is economically and statistically significant in the U.S. stock market as well as crucially related to its information environment. We believe that this is an important contribution to the literature, one that bears important implications for future research on the process of price formation in financial markets.

8. Appendix

Proof of Proposition 1. The basic economy of Section 2.1 nests in the more general setting of Pasquariello (2007). In addition, the distributional assumptions of Section 2.1 imply that $\Sigma_\delta = \rho\Sigma_v$ and $\Sigma_c = \rho^2\Sigma_v = \rho\Sigma_\delta$.

Hence, the linear equilibrium of Proposition 1 follows from Proposition 1 and Remark 1 in Pasquariello (2007). Uniqueness of that equilibrium then ensues from the assumption that $\Sigma_z = \sigma_z^2 I$ (Caballé and Krishnan, 1994, Proposition 3.2). ■

Proof of Remark 1. The statement of the remark follows from the definition of Λ in Proposition 1 (Equation (4)) and the assumption that Σ_v is SPD. Specifically, the latter implies that $\Sigma_v^{1/2} = C\Delta C$ where Δ is a diagonal matrix whose diagonal terms are given by the square roots of the characteristic roots of Σ_v ($\lambda_n > 0$) and C is a matrix whose columns are made of the corresponding orthogonal characteristic vectors c_n, that is, such that $\Sigma_v C = C\Delta$ (Greene, 1997, pp. 36–43). It then follows that $\Sigma_v^{1/2}(l,j) = \sum_{n=1}^{N} c_{nl} c_{jn} \sqrt{\lambda_n}$ will be different from zero if so is $\Sigma_v(l,j)$, and may be so although $\Sigma_v(l,j) = 0$. ■

Proof of Corollary 1. Direct and cross-price impact are decreasing in the number of speculators, since it can be shown that the finite difference $\Delta\left|\Lambda(n,j)\right| = \left|\Lambda(n,j)(\text{at } M+1)\right| - \left|\Lambda(n,j)(\text{at } M)\right| = \left\{\frac{[2+(M-1)\rho]\sqrt{(M+1)\rho} - (2+M\rho)\sqrt{M\rho}}{(2+M\rho)[2+(M-1)\rho]\sigma_z}\right\}$ $\left|\Sigma_v^{1/2}(n,j)\right| < 0$ under most parametrizations (i.e., except in the "small" region of $\{M,\rho\}$ where M is a "small" integer, if speculators' private signals of v are "reasonably" precise). Moreover, $\lim_{M\to\infty}\left|\Lambda(n,j)\right| = 0$. The second part of the statement follows from the fact that $\frac{\partial\left|\Lambda(n,j)\right|}{\partial\rho} = \frac{M[2-(M-1)\rho]}{2\sqrt{M\rho}[2+(M-1)\rho]^2\sigma_z}\left|\Sigma_v^{1/2}(n,j)\right| \geq 0$ if $M \leq \frac{2+\rho}{\rho}$ (i.e., in the presence of "few" speculators with "reasonably" precise private signals of v), and negative otherwise. ■

References

Acharya, V. and Pedersen, L. (2005) Asset pricing with liquidity risk, *Journal of Financial Economics* **77**, 385–410.

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

Amihud, Y. and Mendelson, H. (1986) Asset pricing and the bid-ask spread, *Journal of Financial Economics* **17**, 223–249.

Andrade, S., Chang, C., and Seasholes, M. (2008) Trading imbalances, predictable reversals, and cross-stock effects, *Journal of Financial Economics* **88**, 406–423.

Back, K., Cao, H., and Willard, G. (2000) Imperfect competition among informed traders, *Journal of Finance* **55**, 2117–2155.

Barberis, N., Shleifer, A., and Wurgler, J. (2005) Comovement, *Journal of Financial Economics* **75**, 283–317.

Bennett, P. and Wei, L. (2006) Market structure, fragmentation, and market quality, *Journal of Financial Markets* **9**, 49–78.

Bernanke, B. and Kuttner, K. (2005) What explains the stock market's reaction to Federal Reserve policy?, *Journal of Finance* **60**, 1221–1257.

Bernhardt, D. and Taub, B. (2008) Cross-asset speculation in stock markets, *Journal of Finance* **63**, 2385–2427.

Bessembinder, H. (1999) Trade execution costs on NASDAQ and NYSE: a post-reform comparison, *Journal of Financial and Quantitative Analysis* **34**, 387–407.

Bessembinder, H. (2003a) Issues in assessing trade execution costs, *Journal of Financial Markets* **6**, 233–257.

Bessembinder, H. (2003b) Trade execution costs and market quality after decimalization, *Journal of Financial and Quantitative Analysis* **38**, 747–777.

Blume, M. and Goldstein, M. (1997) Quotes, order flow, and price discovery, *Journal of Finance* **52**, 221–244.

Boehmer, E. and Wu, J. (2008) Order flow and prices, working paper, Texas A&M University.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009) Expected stock returns and variance risk premia, *Review of Financial Studies* **22**, 4463–4492.

Boulatov, A., Hendershott, T., and Livdan, D. (2013) Informed trading and portfolio returns, *Review of Economic Studies* **80**, 35–72.

Brennan, M., Chordia, T., and Subrahmanyam, A. (1998) Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, *Journal of Financial Economics* **49**, 345–373.

Brennan, M. and Subrahmanyam, A. (1995) Investment analysis and price formation in securities markets, *Journal of Financial Economics* **38**, 361–381.

Brennan, M. and Subrahmanyam, A. (1996) Market microstructure and asset pricing: on the compensation for illiquidity in stock returns, *Journal of Financial Economics* **41**, 441–464.

Caballé, J. and Krishnan, M. (1994) Imperfect competition in a multi-security market with risk neutrality, *Econometrica* **62**, 695–704.

Chakrabarty, B. and Moulton, P. (2012) Earnings announcements and attention constraints: the role of market design, *Journal of Accounting and Economics* **53**, 612–634.

Chan, K. (1993) Imperfect information and cross-autocorrelation among stock prices, *Journal of Finance* **48**, 1211–1230.

Chordia, T., Huh, S., and Subrahmanyam, A. (2007) The cross-section of expected trading activity, *Review of Financial Studies* **20**, 709–740.

Chordia, T., Roll, R., and Subrahmanyam, A. (2000) Commonality in liquidity, *Journal of Financial Economics* **56**, 3–28.

Chordia, T., Sarkar, A., and Subrahmanyam, A. (2005) An empirical analysis of stock and bond market liquidity, *Review of Financial Studies* **18**, 85–130.

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 Subrahmanyam, A. (2004) Order imbalance and individual stock returns: theory and evidence, *Journal of Financial Economics* **72**, 485–518.

Comerton-Forde, C., Hendershott, T., Jones, C., Moulton, P., and Seasholes, M. (2010) Time variation in liquidity: the role of market-maker inventories and revenues, *Journal of Finance* **65**, 295–331.

Constantinides, G. (1986) Capital market equilibrium with transaction costs, *Journal of Political Economy* **94**, 842–862.

Corwin, S. (1999) Differences in trading behavior across NYSE specialist firms, *Journal of Finance* **54**, 721–745.

Corwin, S. (2004) Specialist performance and new listing allocations on the NYSE: an empirical analysis, *Journal of Financial Markets* **7**, 27–51.

Coughenour, J. and Deli, D. (2002) Liquidity provision and the organizational form of NYSE specialist firms, *Journal of Finance* **57**, 841–869.

Coughenour, J. and Saad, M. (2004) Common market makers and commonality in liquidity, *Journal of Financial Economics* **73**, 37–69.

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. (1997) Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance* **52**, 1035–1058.

DeLong, B., Shleifer, A., Summers, L., and Waldmann, R. (1990) Noise trader risk in financial markets, *Journal of Political Economy* **98**, 703–738.

Diether, K., Malloy, C., and Scherbina, A. (2002) Differences of opinion and the cross-section of stock returns, *Journal of Finance* **57**, 2113–2141.

Duffie, D., Garleanu, N., and Pedersen, L. (2005) Over-the-counter markets, *Econometrica* **73**, 1815–1847.

Easley, D. and O'Hara, M. (1992) Time and the process of security price adjustment, *Journal of Finance* **47**, 577–604.

Fama, E. and French, K. (1988) Permanent and temporary components of stock prices, *Journal of Political Economy* **96**, 246–273.

Fama, E. and French, K. (1993) Common risk factors in the returns of stocks and bonds, *Journal of Financial Economics* **33**, 3–56.

Foster, F. and Viswanathan, S. (1996) Strategic trading when agents forecast the forecasts of others, *Journal of Finance* **51**, 1437–1478.

Greene, W. (1997) *Econometric Analysis*, Prentice Hall, Upper Saddle River, NJ.

Greenwood, R. (2005) Short- and long-term demand curves for stocks: theory and evidence on the dynamics of arbitrage, *Journal of Financial Economics* **75**, 607–649.

Grinblatt, M. and Titman, S. (1989) Portfolio performance evaluation: old issues and new insights, *Review of Financial Studies* **2**, 393–422.

Grossman, S. and Miller, M. (1988) Liquidity and market structure, *Journal of Finance* **43**, 617–637.

Hartford, J. and Kaul, A. (2005) Correlated order flow: pervasiveness, sources, and pricing effects, *Journal of Financial and Quantitative Analysis* **40**, 29–55.

Hasbrouck, J. (1988) Trades, quotes, inventories, and information, *Journal of Financial Economics* **22**, 229–252.

Hasbrouck, J. (1991) Measuring the information content of stock trades, *Journal of Finance* **46**, 179–207.

Hasbrouck, J. (2002) Stalking the "efficient price" in market microstructure specifications: an overview, *Journal of Financial Markets* **5**, 329–339.

Hasbrouck, J. (2007) *Empirical Market Microstructure*, Oxford University Press, New York.

Hasbrouck, J. and Seppi, D. (2001) Common factors in prices, order flows, and liquidity, *Journal of Financial Economics* **59**, 383–411.

Hendershott, T. and Jones, C. (2005) Island goes dark: transparency, fragmentation, and regulation, *Review of Financial Studies* **18**, 743–793.

Holden, G. and Subrahmanyam, A. (1992) Long-lived private information and imperfect competition, *Journal of Finance* **47**, 247–270.

Huang, M. (2003) Liquidity shocks and equilibrium liquidity premia, *Journal of Economic Theory* **109**, 104–129.
Jones, C., Kaul, G., and Lipson, M. (1994) Transactions, volume, and volatility, *Review of Financial Studies* **7**, 631–651.
Juergens, J. and Lindsey, L. (2009) Getting out early: an analysis of market making activity at the recommending analyst's firm, *Journal of Finance* **64**, 2327–2359.
Kallberg, J. and Pasquariello, P. (2008) Time-series and cross-sectional excess comovement in stock indexes, *Journal of Empirical Finance* **15**, 481–502.
King, M. and Wadhwani, S. (1990) Transmission of volatility between stock markets, *Review of Financial Studies* **3**, 5–33.
Kodres, L. and Pritsker, M. (2002) A rational expectations model of financial contagion, *Journal of Finance* **57**, 769–799.
Kyle, A. (1985) Continuous auctions and insider trading, *Econometrica* **53**, 1315–1335.
Kyle, A. and Xiong, W. (2001) Contagion as a wealth effect, *Journal of Finance* **56**, 1401–1440.
Lee, C. and Ready, M. (1991) Inferring trade direction from intraday data, *Journal of Finance* **46**, 733–746.
O'Hara, M. and Ye, M. (2011) Is market fragmentation harming market quality?, *Journal of Financial Economics* **100**, 459–474.
Pasquariello, P. (2007) Imperfect competition, information heterogeneity, and financial contagion, *Review of Financial Studies* **20**, 391–426.
Pasquariello, P. and Vega, C. (2007) Informed and strategic order flow in the bond markets, *Review of Financial Studies* **20**, 1975–2019.
Pasquariello, P. and Vega, C. (2009) The on-the-run *liquidity* phenomenon, *Journal of Financial Economics* **92**, 1–24.
Pastor, L. and Stambaugh, R. (2003) Liquidity risk and expected stock returns, *Journal of Political Economy* **113**, 642–685.
Qu, S., Starks, L., and Yan, H. (2004) Risk, dispersion of analyst forecasts, and stock returns, working paper, University of Texas.
Sadka, R. and Scherbina, A. (2007) Analyst disagreement, mispricing, and liquidity, *Journal of Finance* **62**, 2367–2403.
Shiller, R. (1989) Comovements in stock prices and comovements in dividends, *Journal of Finance* **44**, 719–729.
Shleifer, A. (2000) *Inefficient markets*, Oxford University Press, New York.
Stoll, H. (2001) Market fragmentation, *Financial Analysts Journal* **57**, 16–20.
Tookes, H. (2008) Information, trading, and product market interactions: cross-sectional implications of informed trading, *Journal of Finance* **63**, 379–413.
Vayanos, D. (1998) Transaction costs and asset prices: a dynamic equilibrium model, *Review of Financial Studies* **11**, 1–58.
Vega, C. (2006) Stock price reaction to public and private information, *Journal of Financial Economics* **82**, 103–133.
Veldkamp, L. (2006) Information markets and the comovement of asset prices, *Review of Economic Studies* **73**, 823–845.
Watanabe, M. (2008) A model of stochastic liquidity, working paper, Rice University.
Wermers, R. (1999) Mutual fund trading and the impact on stock prices, *Journal of Finance* **54**, 581–622.
Yu, J. (2011) Disagreement and return predictability of stock portfolios, *Journal of Financial Economics* **99**, 162–183.