ELSEVIER

Journal of Financial Markets 9 (2006) 246–273

Journal of FINANCIAL MARKETS

www.elsevier.com/locate/finmar

The impact of preferencing on execution quality☆

Chen He, Elizabeth Odders-White*, Mark J. Ready

School of Business, University of Wisconsin, Madison, WI 53706, USA

Available online 2 May 2006

Abstract

We examine the impact of preferencing on execution quality for NASDAQ and NYSE-listed stocks. Our theoretical model demonstrates that realized spreads are more reliable than effective spreads in the presence of preferencing, but even realized spreads are a poor measure of execution quality if the stocks being compared have different degrees of information asymmetry. We provide a new measure of the costs of preferencing that is independent of asymmetric information. Using data from the SEC 11Ac1-5 reports for marketable orders of up to 2000 shares, we find that both realized spreads and our preferencing measure are lower for NYSE-listed stocks.

JEL classification: G10; D43

Keywords: Market microstructure; Preferencing; Execution quality; Spreads

1. Introduction

Stocks in the U.S. trade not only on their primary exchange, but also in other markets. While alternative market centers can compete for order flow by posting aggressive quotes, they also obtain orders through preferencing agreements. Preferencing agreements between market centers and brokers who route customers' orders allow market centers to obtain order flow regardless of their quotes. Although these agreements are fairly widespread for both New York Stock Exchange (NYSE) listed and NASDAQ stocks, they are much more prevalent for NASDAQ stocks. Accordingly, to estimate the impact of preferencing on

☆We thank David Brown, Kenneth Kavajecz, and Masako Ueda for helpful comments. We gratefully acknowledge financial support from the Graduate School and the Puelicher Center for Banking Education at the University of Wisconsin, Madison.

*Corresponding author. Tel.: +1 608 263 1254; fax: +1 608 265 4195.

1386-4181/$ - see front matter
doi:10.1016/j.finmar.2006.02.003

execution quality, we compare various measures of execution costs across these two markets.

Preferencing agreements can take various forms. In some cases, the preferencing broker agrees to send customer orders to the preferenced market center, which in turn agrees to follow predetermined procedures for setting the execution price, such as agreeing to match the national best bid or offer. The preferenced market center may also pay the preferencing broker for the routed orders. In another case, called *internalization*, the preferenced market center and the broker are part of the same corporation. Both theory and anecdotal evidence suggest that market makers try to use preferencing to attract orders from investors whose trading decisions are not based on short-term price information.

Proponents of preferencing argue that as long as preferenced orders are given reasonable prices, customers are not disadvantaged. In fact, the U.S. Securities and Exchange Commission's 1997 ''Report on the Practice of Preferencing'' finds effective spreads are no higher at preferenced market centers, and thus concludes that preferencing does not diminish market quality. Battalio (1997), Hansch et al. (1999), and Peterson and Sirri (2003) have examined effective spreads and also find that preferencing has no clear negative impact on execution costs. Others assert that preferencing reduces dealers' incentives to offer competitive execution and limits the ability of orders to interact, both of which may harm investors. For example, Huang and Stoll (1996) and Chung et al. (2004), among others, document a positive relationship between preferencing and execution costs.

To guide our analysis, we extend the Kyle (1985) model to include preferencing of a portion of the uninformed order flow. Our model shows that preferenced market makers profit at the expense of both uninformed and informed investors even though the preferenced market maker matches the effective spreads given to non-preferenced uninformed orders. As the fraction of order flow that is preferenced increases, uninformed traders in both the primary and preferenced markets pay wider spreads, and the informed trader scales back trading, thereby reducing profits. More than half of the preferenced market maker's gain from preferencing comes at the expense of the uninformed traders.

Our model also shows that differences in effective spreads do not fully reflect the negative impact of preferencing on investors because they capture not merely dealers' profits, but also the profits of informed traders. Because the increase in dealer profits is partially offset by a decline in informed traders' profits, effective spreads underestimate the true costs of preferencing. Realized spreads, in contrast, reflect only dealer profits and thus capture the combined cost of preferencing borne by the uninformed and informed traders. This suggests that realized spreads are a better measure of market quality than effective spreads in the presence of different degrees of preferencing. Still, realized and effective spreads are equally sensitive in percentage terms to differences in information asymmetry when preferencing dealers match the effective spread paid by uninformed traders in the competitive portion of the market. This implies that *both* realized and effective spreads are poor measures of execution quality if the stocks that are compared have differences in asymmetric information.

Our model yields a new measure of preferencing that, unlike effective and realized spreads, completely controls for the level of information asymmetry. This makes our measure useful for comparisons across stocks and across exchanges. Because dealer profits increase with the degree of preferencing, this new measure provides an indication of the extent of the costs preferencing imposes on traders.

Guided by our model, we compare market quality for NASDAQ and NYSE-listed stocks using monthly execution quality reports published by all market centers pursuant to U.S. Securities and Exchange Commission (SEC) (2001) Rule 11Ac1-5.[1] The SEC's definition of "market center" encompasses NASDAQ dealers, NYSE or regional specialists, electronic communications networks (ECNs), and alternative trading systems. The 11Ac1-5 reports include share-weighted average effective and realized spreads, which can be used to compute our preferencing measure, along with time to execution and execution rates. Quoted national best bid or offer (NBBO) spreads can be computed from these data, as well.

Our study period spans January through June 2004. We examine the combination of market and marketable limit orders (*marketable orders*), and use regression analysis to control for differences in firm characteristics. We focus on the two order size categories below 2000 shares, which are most likely to be covered by preferencing agreements, and we run separate tests for four different firm-size categories.

We find that both effective and realized spreads are wider for NASDAQ stocks than for NYSE-listed stocks in all firm-size categories. The fact that the differences are larger for realized spreads suggests that the poorer execution quality for NASDAQ stocks may be related to a greater degree of preferencing. The higher preferencing measures for NASDAQ stocks confirm this hypothesis, and suggest that dealers in some NASDAQ stocks capture additional profits of over 50% of the effective spread. When we examine the results by market center, we find that, consistent with the model, differences in realized spreads are more pronounced than differences in effective spreads, and preferenced market centers generally have higher realized spreads than non-preferenced market centers.

2. The model

Our extension of the single-period Kyle (1985) model serves three purposes. First, it allows us to examine the strategic response of the informed trader to preferencing. Second, it demonstrates the potential impact of preferencing-induced fragmentation on effective and realized spreads, which helps to guide our empirical analysis. Finally, it yields a new measure of the costs of preferencing that controls completely for the degree of information asymmetry. Together, these three points distinguish our model from other work in this area.[2]

Chakravarty and Sarkar (2002) use a modified Kyle model to examine dual trading. They also consider preferencing in the context of their model, but there are important distinctions between our model and theirs. In the Chakravarty and Sarkar model, the preferencing broker is a "simultaneous dual trader," who generates a proprietary order that partially offsets his uninformed customer's order and then sends the net of these two orders to a competitive market maker. The competitive market maker then sets the execution price based on the aggregate order flow, which also includes orders from multiple informed traders and from a "piggybacking" broker who has seen the informed traders' orders. In our model, a separate market center executes the preferenced

[1] Huang and Stoll (1996), Bessembinder and Kaufman (1997), Bessembinder (1999), and Weston (2000) conduct cross-exchange comparisons of execution quality in the pre-decimal period using trade and quote data.

[2] See, e.g., Chordia and Subrahmanyam (1995), Easley et al. (1996), Kandel and Marx (1999), Battalio and Holden (2001), Chakravarty and Sarkar (2002), Parlour and Rajan (2003), and Parlour and Seppi (2003).

uninformed order, which is much closer to the practice in U.S. markets. In addition, our model examines gross rather than net order flow, which allows us to examine trading costs on a per-share basis, so we can use our model to interpret the 11Ac1-5 reports.

2.1. Execution quality without preferencing

Following Kyle (1985), our model includes three types of traders: liquidity traders whose net share demand is random and endogenously determined; a single risk-neutral informed trader who submits orders based on a signal about the true value of the asset; and a risk-neutral, competitive (zero expected profit) market maker who clears the combined orders from the uninformed and informed traders. The Kyle framework is attractive because it lets us examine the strategic behavior of the informed trader, and separately examine the profits of both the informed and uninformed traders. It does not explicitly include spreads, however, so we expand the model to include a measure of gross order flow and to develop expressions for share-weighted average effective and realized spreads per share. These share-weighted average spreads are included in the 11Ac1-5 statistics, which we describe in detail in Section 3 and analyze in Section 4. The term "share-weighted" is used in the rule to distinguish the statistics from order-weighted averages. For example, suppose a market center executed two orders during the month — a 100-share order with an effective spread of \$0.10 per share and a 400-share order with an effective spread of \$0.20 per share. The share-weighted average effective spread is $(100*0.10+400*0.20)/500 = \$0.18$ per share, compared to an order-weighted average of $(0.10+0.20)/2 = \$0.15$ per share.

For our model, we adopt Kyle's notation:

p_0 = expected value per share of the asset based on publicly available information (prior to observing the order flow)

$\tilde{v}$ = informed trader's signal of the true value of the asset per share, which is normally distributed with expected value equal to p_0

Σ_0 = variance of $\tilde{v}$

$\tilde{u}$ = uninformed traders' net order flow (buy orders – sell orders) in shares, which is normally distributed with an expected value of 0

σ_u^2 = variance of $\tilde{u}$

$\tilde{x} = X(\tilde{v})$ = number of shares submitted by the informed trader (negative means the informed trader is selling), and

$P(\tilde{x}+\tilde{u})$ = price set by the market maker to execute $\tilde{x}+\tilde{u}$ shares, which is the total net order flow from the informed and uninformed traders.

Kyle shows that the linear equilibrium to this game is given by

$$\tilde{x} = X(\tilde{v}) = \beta(\tilde{v}-p_0) \quad \text{and} \quad P(\tilde{x}+\tilde{u}) = p_0 + \lambda(\tilde{x}+\tilde{u})$$

where $\beta = (\sigma_u^2/\Sigma_0)^{1/2}$ and $\lambda = (1/2)\,(\sigma_u^2/\Sigma_0)^{-1/2}$. For the derivations below, it is useful to note that $\mathrm{Var}[\tilde{x}] = \beta^2\mathrm{Var}[\tilde{v}] = (\sigma_u^2/\Sigma_0)\Sigma_0 = \sigma_u^2$.

Kyle's model focuses on *net* order flow, so to allow us to characterize average effective and realized spreads per share we need to make an additional assumption about the uninformed trading that allows us to define *gross* order flow (total buys plus total sells). We assume there are a large number of uninformed investors. In each period, a fixed number, N_u, of these investors receive an exogenous liquidity shock that causes them to need to trade. Each investor's order $\tilde{u}_i$ is an independent and identically distributed normal random variable with mean zero and standard deviation σ_n. A positive (negative) value of $\tilde{u}_i$ means that uninformed investor i is buying (selling). Total net order flow from the uninformed traders is given by

$$\tilde{u} = \sum_{i=1}^{N_u} \tilde{u}_i,$$

which is normally distributed with mean 0 and standard deviation, $\sigma_u = \sigma_n \sqrt{N_u}$.

The total net order flow is $\tilde{x} + \tilde{u}$. The total *gross* order flow is

$$\tilde{G} = \sum_{i=1}^{N_u} |\tilde{u}_i| + |\tilde{x}|.$$

We can also define gross buy orders as

$$\tilde{B} = \sum_{i=1}^{N_u} \max[\tilde{u}_i, 0] + \max[\tilde{x}, 0]$$

and gross sell orders as $\tilde{S} = \tilde{G} - \tilde{B}$. With these definitions, the net order flow, $\tilde{B} - \tilde{S}$, can be shown to equal $\tilde{x} + \tilde{u}$, as required.[3] For a normal random variable with mean zero and variance σ^2 the expectation of the absolute value is $\sigma\sqrt{2/\pi}$. This means that the expected value of the gross order flow is given by

$$\mathrm{E}[\tilde{G}] = \sqrt{2/\pi}[N_u \sigma_n + \sigma_x] = \sigma_u \sqrt{2/\pi}\left[1 + \sqrt{N_u}\right]$$

Although our model explicitly includes gross order flow, we continue to assume that the market maker's price is a function of only the net order flow, so the informed trader's strategy and the market maker's pricing function are the same as those given by Kyle. The fact that the informed trader can make gross order flow uninformative by submitting orders on both sides of the market suggests that net order flow is a logical input to the pricing function.

In many respects, our model is similar to the NYSE's procedure at the open, where buy and sell market orders are first paired off, and the specialist (in combination with the limit orders on the book) then trades against the remaining imbalance, with all orders receiving the same opening price. Similar procedures are used after trading halts. The NYSE's protocol during the regular trading day is somewhat different, but one of the key elements of the NYSE's auction is the netting of market orders whenever possible.

In our model, everyone trades at the same price, which equals $p_0 + \lambda(\tilde{x} + \tilde{u})$. The realized spread is defined by comparing trade prices to a subsequent reference price, and in the case of SEC Rule 11Ac1-5, this subsequent reference price is the quote midpoint five minutes after the trade. In our model, one candidate for this subsequent reference price is the execution price for the trade, because, as shown by Kyle, the execution price is the conditional expected value of $\tilde{v}$ given the net order flow. According to this definition, the realized spread is always exactly zero.

[3]See "Result 1" in Appendix A.

Alternatively, one could assume that $\tilde{v}$ is revealed immediately after the trade, so that $\tilde{v}$ is the subsequent reference price. In this case, the aggregate realized (half) spread is given by:[4]

$$\begin{aligned}\tilde{B}([p_0+\lambda(\tilde{x}+\tilde{u})]-\tilde{v})+\tilde{S}(\tilde{v}-[p_0+\lambda(\tilde{x}+\tilde{u})]) &= (\tilde{B}-\tilde{S})([p_0+\lambda(\tilde{x}+\tilde{u})]-\tilde{v})\\ &= (\tilde{x}+\tilde{u})(p_0+\lambda(\tilde{x}+\tilde{u})-\tilde{v}).\end{aligned}$$

Although this expression is random, it can be shown to have an expected value of zero by using the law of iterated expectations in combination with Kyle's result that $\mathrm{E}[p_0+\lambda(\tilde{x}+\tilde{u})-\tilde{v}|\tilde{x}+\tilde{u}]=0$.[5] The result that the expected realized spread equals zero is intuitive, because we are aggregating across all traders. The market maker earns zero expected profit, so the total combined expected profit earned by informed and uninformed traders is also zero. In our analysis, we choose $\tilde{v}$ as the subsequent reference price for calculating realized spreads.

The effective spread compares the trade price to the quote midpoint at the time the order arrived, and in our model, p_0 is the natural choice to represent this pre-trade quote midpoint. Accordingly, the aggregate effective (half) spread is given by

$$\tilde{B}([p_0+\lambda(\tilde{x}+\tilde{u})]-p_0)+\tilde{S}(p_0-[p_0+\lambda(\tilde{x}+\tilde{u})])=(\tilde{B}-\tilde{S})\lambda(\tilde{x}+\tilde{u})=\lambda(\tilde{x}+\tilde{u})^2.$$

The informed order flow, $\tilde{x}$, and the uninformed order flow, $\tilde{u}$, both have zero mean, variance σ_u^2, and are independent of each other. Thus, the expected value of the aggregate effective (half) spread is $\lambda 2\sigma_u^2=\sigma_u\sqrt{\Sigma_0}$.

We ultimately want to use our model to analyze the available data. The execution cost statistics used in our empirical tests below are share-weighted six-month averages, so they represent many repetitions of the orders and price responses in our model. Letting t denote each period, the share-weighted average effective spread across T repetitions of our model is given by

$$\sum_{t=1}^{T}\left(\frac{\tilde{G}_t}{\sum_{t=1}^{T}\tilde{G}_t}\right)\left(\frac{\lambda(\tilde{x}_t+\tilde{u}_t)^2}{\tilde{G}_t}\right)=\frac{(1/T)\sum_{t=1}^{T}\lambda(\tilde{x}_t+\tilde{u}_t)^2}{(1/T)\sum_{t=1}^{T}\tilde{G}_t}.$$

As T gets large, both the denominator and the numerator in the above expression converge to their expected values, so the share-weighted average effective spread converges to the ratio of the expected value of the aggregate effective (half) spread to the expected value of the gross order flow. This ratio is equal to

$$\frac{\mathrm{E}\left[\lambda(\tilde{x}+\tilde{u})^2\right]}{\mathrm{E}\left[\tilde{G}\right]}=\frac{\sqrt{\pi\Sigma_0}}{\sqrt{2}\left(1+\sqrt{N_u}\right)}.$$

Our statistical tests use sample estimates of the numerator and denominator, so sampling error in the denominator may bias our estimate of the ratio. We address this issue in Section 4 and in Appendix B.

The profit earned by the informed trader is his or her volume times the difference between the true value and the trade price. Thus, the expected profit is

$$\mathrm{E}[\tilde{x}(\tilde{v}-[p_0+\lambda(\tilde{x}+\tilde{u})])]=\mathrm{E}[\tilde{x}(\tilde{v}-p_0)]-\mathrm{E}[\lambda\tilde{x}^2]-\mathrm{E}[\lambda\tilde{x}\tilde{u}].$$

[4] In some applications, such as the 11Ac1-5 reports, this value would be multiplied by two to make the value comparable to the quoted spread. We use the label "(half)" to clarify that we are not multiplying by two.

[5] See "Result 2" in Appendix A.

Given the definitions of $\tilde{x}$ and λ, along with the fact that $\tilde{x}$ and $\tilde{u}$ are independent, the expression for the expected profit of the informed trader reduces to $(1/2)\sigma_u\sqrt{\Sigma_0}$. The expected aggregate (half) spread paid by the informed trader is $\mathrm{E}[\tilde{x}([p_0 + \lambda(\tilde{x} + \tilde{u})] - p_0)] = \mathrm{E}[\lambda\tilde{x}^2] + \mathrm{E}[\lambda\tilde{x}\tilde{u}]$. As mentioned above, $\mathrm{Var}[\tilde{x}] = \sigma_u^2$, so this also reduces to $(1/2)\sigma_u\sqrt{\Sigma_0}$. The expected profit and the expected aggregate (half) spread are equal because it is optimal for the informed trader to choose an order size that drives the price half way (in expected value terms) from p_0 to $\tilde{v}$.

The expected aggregate (half) spread paid by the uninformed traders is

$$\mathrm{E}[\tilde{u}([p_0 + \lambda(\tilde{x} + \tilde{u})] - p_0)] = \mathrm{E}[\lambda\tilde{u}^2] + \mathrm{E}[\lambda\tilde{x}\tilde{u}] = \lambda\sigma_u^2 = (1/2)\sigma_u\sqrt{\Sigma_0}.$$

The uninformed traders' expected loss also reduces to $(1/2)\sigma_u\sqrt{\Sigma_0}$ because it is simply equal to the informed trader's expected profit. Although the uninformed traders' effective aggregate (half) spread is computed by comparing the expected trade price to p_0, and the expected loss is computed by comparing the expected trade price to $\tilde{v}$, these two quantities are equal because the uninformed order flow is uncorrelated with the difference between p_0 and $\tilde{v}$, so $\mathrm{E}[\tilde{v}|\tilde{u}] = p_0$.

Informed and uninformed traders pay the same average aggregate effective (half) spread and have the same expected net order flow. Expected gross order flow is larger for the uninformed traders (when $N_u > 1$), so the uninformed traders pay a lower share-weighted average effective (half) spread than the informed traders. The long-run share-weighted average effective (half) spread per share paid by uninformed traders is

$$\frac{\sqrt{\pi\Sigma_0}}{2\sqrt{2N_u}}$$

(i.e., their expected aggregate effective (half) spread divided by their expected gross volume). Intuitively, the lower spread for uninformed orders reflects the netting of uninformed buy and sell orders.[6]

2.2. *Execution quality with preferencing*

We now add two additional players to our model: a broker who controls a fraction f of the uninformed order flow (orders from fN_u of the uninformed traders), and a market maker who enters a preferencing agreement with the broker. The broker agrees to route all orders to the preferenced market maker, who in turn agrees to execute the orders and match the long-run average effective spread for the *uninformed orders* in the competitive portion of market. The remaining $(1 - f)$ of the uninformed orders that are not controlled by the broker are routed to the competitive portion of market, as are the orders from the informed trader.

The two new players in our model do not explicitly behave strategically, but their presence does affect the outcome in the competitive portion of the market. Recall that, under the informed trader's optimal strategy, the variance of the informed trader's order flow is equal to the variance of the net uninformed order flow. The variance of the net uninformed order flow in the primary market is lower with preferencing (equal to

[6] Jones and Lipson (2003) present evidence that retail orders do in fact receive lower effective spreads on the NYSE, which is consistent with our model, but may also reflect some degree of sorting uninformed orders from informed orders inside the NYSE (see Benveniste et al. (1992)).

$(1-f)\sigma_u^2$), so the informed trader trades less aggressively, with $\sigma_x = \sqrt{(1-f)}\sigma_u$. The competitive market maker understands this and reacts accordingly when establishing the clearing price.

The new equilibrium in the competitive portion of the market has

$$\beta^p = [(1-f)\sigma_u^2/\Sigma_0]^{1/2} = \sqrt{(1-f)}\beta$$

and

$$\lambda^p = (1/2)[(1-f)\sigma_u^2/\Sigma_0]^{-1/2} = \lambda/\sqrt{(1-f)}$$

where the superscript p denotes the values with preferencing. Kyle shows that the "efficiency" of the resulting equilibrium price in the competitive portion of the market is not affected by the diversion of some of the uninformed order flow. The variance of the pricing error, $\mathrm{Var}[p_0 + \lambda(\tilde{x}+\tilde{u}) - \tilde{v}]$, is equal to $(1/2)\Sigma_0$, which is independent of the intensity of the uninformed trader.

The expected aggregate effective (half) spread in the competitive portion of the market is $\sqrt{(1-f)}\sigma_u\sqrt{\Sigma_0}$, and the informed trader's expected profit is $(1/2)\sqrt{(1-f)}\sigma_u\sqrt{\Sigma_0}$. The expected gross volume in the competitive portion of the market is

$$\sqrt{2/\pi}[(1-f)N_u\sigma_n + \sigma_x] = \sigma_u\sqrt{2/\pi}\left[(1-f)\sqrt{N_u} + \sqrt{(1-f)}\right],$$

where the right-hand side uses $\sigma_n = \sigma_u/\sqrt{N_u}$ and $\sigma_x = \sqrt{(1-f)}\sigma_u$.

We assume that the agreement between the preferencing broker and the preferenced market center specifies that the share-weighted average effective spread given to the preferenced orders be equal to the share-weighted average for the *uninformed orders* in the competitive portion of the market across many trading sessions (with many information events). In the competitive portion of the market, the uninformed traders' expected aggregate effective (half) spread is $(1/2)\sqrt{(1-f)}\sigma_u\sqrt{\Sigma_0}$, and their expected volume is $\sigma_u\sqrt{2/\pi}\left[(1-f)\sqrt{N_u}\right]$. Accordingly, the long-run share-weighted average effective (half) spread per share for the *uninformed orders* in the competitive portion of the market is $\overline{es}^{p,c,u} = \frac{1}{2}\sqrt{\Sigma_0\pi/2}\Big/\left[\sqrt{(1-f)N_u}\right]$.

The p in the superscript denotes that the entire market includes a preferencing arrangement, the c denotes the competitive portion of the market, and the u denotes uninformed traders. The preferencing agreement specifies that each order be charged $\overline{es}^{p,p} = \overline{es}^{p,c,u}$ per share, and the expected gross volume in the preferenced portion of the market is $\sigma_u\sqrt{2/\pi}f\sqrt{N_u}$. Each share in the preferenced portion of the market is charged $\overline{es}^{p,c,u}$, so the expected aggregate (half) spread in the preferenced portion of the market is simply the expected volume times $\overline{es}^{p,c,u}$, which is

$$\begin{aligned}\sigma_u\sqrt{2/\pi}f\sqrt{N_u}(\overline{es}^{p,c,u}) &= \sigma_u\sqrt{2/\pi}f\sqrt{N_u}\left(\frac{1}{2}\sqrt{\Sigma_0\pi/2}\Big/\left[\sqrt{(1-f)N_u}\right]\right)\\ &= \frac{1}{2}\sigma_u\sqrt{\Sigma_0}\frac{f}{\sqrt{1-f}}.\end{aligned}$$

The expected aggregate effective (half) spread across the entire market is the sum of the amounts from the competitive and preferenced portions, which is equal to

$$\sqrt{(1-f)}\sigma_u\sqrt{\Sigma_0} + \frac{1}{2}\sigma_u\sqrt{\Sigma_0}\frac{f}{\sqrt{1-f}} = \sigma_u\sqrt{\Sigma_0}\frac{1-\frac{1}{2}f}{\sqrt{1-f}}.$$

Dividing by expected total gross volume (the expected gross volume across the entire market is $\sigma_u\sqrt{2/\pi}\left[\sqrt{N_u}+\sqrt{(1-f)}\right]$) gives the long-run share-weighted average across the entire market, which we denote as $\overline{es}^p$

$$\overline{es}^p=\frac{\sigma_u\sqrt{\Sigma_0}\left(1-\frac{1}{2}f\right)/\sqrt{(1-f)}}{\sigma_u\sqrt{2/\pi}\left[\sqrt{N_u}+\sqrt{(1-f)}\right]}=\frac{\sqrt{\Sigma_0\pi/2}\left(1-\frac{1}{2}f\right)}{\sqrt{N_u(1-f)}+(1-f)}. \tag{1}$$

The expected aggregate realized (half) spread in the competitive portion of the market is still zero. The preferenced market maker charges buyers $p_0+\overline{es}^{p,c,u}$ and pays sellers $p_0-\overline{es}^{p,c,u}$, so the aggregate realized (half) spread in the preferenced portion of the market is:

$$\sum_{i=1}^{fN_u}\left[\max(u_i,0)(p_0+\overline{es}^{p,c,u}-\tilde{v})+\max(-u_i,0)(\tilde{v}-(p_0-\overline{es}^{p,c,u}))\right]$$
$$=\overline{es}^{p,c,u}\sum_{i=1}^{fN_u}|\tilde{u}_i|+(p_0-\tilde{v})\sum_{i=1}^{fN_u}u_i.$$

The true value $\tilde{v}$ and the uninformed order flow are independent, and $p_0=\mathrm{E}[\tilde{v}]$, so the expectation of the second term on the right-hand side of the expression is zero. Again, because the expectation of the absolute value of a normal random variable with mean zero and variance σ^2 is $\sigma\sqrt{2/\pi}$, the expected aggregate realized (half) spread in the preferenced portion of the market (and across the market as a whole) is

$$\overline{es}^{p,c,u}\sqrt{2/\pi}fN_u\sigma_n=\overline{es}^{p,c,u}\sqrt{2/\pi}f\sqrt{N_u}\sigma_u=\frac{1}{2}f\sigma_u\sqrt{\Sigma_0}\Big/\sqrt{(1-f)}.$$

The expected gross volume across the entire market is $\sigma_u\sqrt{2/\pi}\left[\sqrt{N_u}+\sqrt{(1-f)}\right]$, so the long-run share-weighted average realized (half) spread per share is

$$\overline{rs}^p=\frac{\frac{1}{2}f\sigma_u\sqrt{\Sigma_0}/\sqrt{(1-f)}}{\sigma_u\sqrt{2/\pi}\left[\sqrt{N_u}+\sqrt{(1-f)}\right]}=\frac{\frac{1}{2}f\sqrt{\Sigma_0\pi/2}}{\sqrt{N_u(1-f)}+(1-f)}. \tag{2}$$

Combining (1) and (2) gives

$$\overline{rs}^p=\frac{f}{(2-f)}\overline{es}^p. \tag{3}$$

2.3. Sensitivity of measures to preferencing and adverse selection

To examine the impact of preferencing on execution costs, we differentiate expressions (1) and (3) with respect to f, yielding the expressions:

$$\frac{\mathrm{d}\overline{es}^p}{\mathrm{d}f}=\sqrt{\Sigma_0\pi/2}\left[\frac{-\frac{1}{2}\left(\sqrt{N_u(1-f)}+(1-f)\right)+\left(1-\frac{1}{2}f\right)\left(\frac{1}{2}\sqrt{N_u/(1-f)}+1\right)}{\left(\sqrt{N_u(1-f)}+(1-f)\right)^2}\right],$$
$$=\frac{1}{2}\sqrt{\Sigma_0\pi/2}\left[\frac{1+\left(\frac{\frac{1}{2}f\sqrt{N_u}}{\sqrt{1-f}}\right)}{\left(\sqrt{N_u(1-f)}+(1-f)\right)^2}\right],$$

and

$$\frac{\mathrm{d}\overline{rs}^p}{\mathrm{d}f} = \frac{2}{(2-f)^2}\overline{es}^p + \frac{f}{2-f}\frac{\mathrm{d}\overline{es}^p}{\mathrm{d}f}.$$

Both are positive, indicating that all else equal, higher preferencing results in higher effective and realized spreads.

By calculating the difference between $\mathrm{d}\overline{rs}^p/\mathrm{d}f$ and $\mathrm{d}\overline{es}^p/\mathrm{d}f$, we can compare the sensitivity of each measure to changes in the degree of preferencing. Subtracting $\mathrm{d}\overline{es}^p/\mathrm{d}f$ from $\mathrm{d}\overline{rs}^p/\mathrm{d}f$ and simplifying yields:

$$\begin{aligned}\frac{\mathrm{d}\overline{rs}^p}{\mathrm{d}f} - \frac{\mathrm{d}\overline{es}^p}{\mathrm{d}f} &= \frac{2}{(2-f)^2}\overline{es}^p + \frac{2(f-1)}{2-f}\frac{\mathrm{d}\overline{es}^p}{\mathrm{d}f}\\ &= \frac{\sqrt{N_u(1-f)\Sigma_0\pi/2}}{2\left(\sqrt{N_u(1-f)}+(1-f)\right)^2}.\end{aligned}$$

This expression is positive, so preferencing increases realized spreads faster than it increases effective spreads. Effective spreads increase more slowly because the increase in dealer profits from preferencing is partially offset by a decline in informed traders' profits. Informed traders suffer because they scale back their trading in response to reduced liquidity in the primary market. Consequently, effective spreads underestimate the true costs that investors bear as a result of preferencing. This suggests that realized spreads are a better measure of execution quality in the presence of preferencing, and implies that results based on effective spreads should be interpreted with caution.

Expressions (1) and (2) also reveal that both effective spreads and realized spreads are affected by the degree of information asymmetry Σ_0 for the stock. The fact that realized spreads depend on adverse selection is contrary to the typical view. Eq. (3) implies that

$$\frac{\mathrm{d}\overline{rs}^p}{\mathrm{d}\Sigma_0} = \frac{f}{2-f}\frac{\mathrm{d}\overline{es}^p}{\mathrm{d}\Sigma_0}.$$

The value of f must be less than or equal to 1, so $f/(2-f)\leqslant 1$, which means effective spreads are more sensitive to Σ_0 in dollar terms. On the other hand, combining these two expressions gives

$$\frac{1}{\overline{es}^p}\frac{\mathrm{d}\overline{es}^p}{\mathrm{d}\Sigma_0} = \frac{1}{\overline{rs}^p}\frac{\mathrm{d}\overline{rs}^p}{\mathrm{d}\Sigma_0},$$

which implies effective and realized spreads are equally sensitive to differences in asymmetric information in percentage terms. Thus, both effective and realized spreads are poor measures of execution quality if the stocks that are compared have different degrees of asymmetric information.

Expressions (1) and (2), along with the comparative statics above, suggest that the ratio of the realized spread to the effective spread provides a natural measure of the degree of preferencing in the market. This ratio, which we call the preferencing measure and denote *pm*:

$$pm = \overline{rs}^p/\overline{es}^p = \frac{f}{2-f} \tag{4}$$

is monotonically increasing in f and is independent of Σ_0, so it does not suffer from the shortcomings of effective and realized spreads. The preferencing measure can be interpreted as the fraction of the effective spread earned by liquidity providers.

Solving (4) directly for f yields an alternative measure of preferencing:

$$pm_a = f = \frac{2pm}{(1+pm)} = \frac{\left(2\overline{rs}^p/\overline{es}^p\right)}{\left(1+\left(\overline{rs}^p/\overline{es}^p\right)\right)}.$$

Although this measure is appealing because it assumes the value of f when $\overline{es}^p$and $\overline{rs}^p$ are known, noise in the estimates of realized and effective spreads for individual stocks can make the measure unstable, even when averaging over six months of data.[7]

2.4. Impact of preferencing on welfare

Our model can be used to examine the impact of preferencing on the informed and uninformed traders and the preferenced market maker. Preferenced market makers obviously benefit from these agreements. They may share their gains with the routing brokers, either through payment for order flow or internalization. (Recall that internalization refers to cases in which the market maker is owned by the brokerage firm.) Thus, for the remainder of this section we use ''broker/market maker'' to refer to the combination of these entities. To the extent that some of the rents transferred from the market maker to the routing broker are in turn transferred to the broker's customers, our estimates of the costs borne by the uninformed traders will be overstated. We do not know of any examples where payments for order flow are explicitly rebated to customers. Also, in Section 4 we show that market making profits appear higher in NASDAQ stocks than in comparable NYSE stocks, but we do not know of any brokers who charge lower commissions for trading NASDAQ stocks.

The broker/market maker expected gains, which are given by the expected aggregate realized (half) spread in the preferenced portion of the market, $\frac{1}{2}f\sigma_u\sqrt{\Sigma_0}/\sqrt{(1-f)}$, must equal the difference between uninformed traders' expected loss and the informed trader's expected profit. Clearly, preferencing reduces the expected profits of the informed trader in our model, because (s)he trades fewer shares and yet the trades move prices by the same amount. The expected profit of the informed trader is equal to $(1/2)\sqrt{(1-f)}\sigma_u\sqrt{\Sigma_0}$, which decreases in f.

Preferencing also harms the uninformed traders in our model.[8] The long-run share-weighted average effective (half) spread per share paid by uninformed traders is the same in both the competitive and preferenced portions of the market. It is equal to $\frac{1}{2}\sqrt{\Sigma_0\pi/2}/\left[\sqrt{(1-f)N_u}\right]$, which is increasing in f.

In fact, more than half of the broker/market maker gain comes at the expense of the uninformed traders. If we denote the broker/market maker profit by *mmp*, the informed trader's profit by *itp*, and the uninformed traders' total loss by *utl*, we have $\mathrm{d}mmp/df = (\mathrm{d}utl/\mathrm{d}f) - (\mathrm{d}itp/\mathrm{d}f)$. By calculating the derivatives of *mmp* and *itp* with

[7] We revisit this issue in Section 4.

[8] In the Chakravarty and Sarkar (2002) model, the uninformed traders can be hurt by preferencing, but only if the presence of preferencing reduces the number of informed traders that operate in the market (see the proof to their proposition 5). Fewer informed traders means that they compete less aggressively with each other, thereby extracting higher oligopoly profit.

respect to f, we can show that $\frac{1}{2}(\mathrm{d}mmp/\mathrm{d}f)+(\mathrm{d}itp/\mathrm{d}f)>0$, which implies that $\mathrm{d}utl/\mathrm{d}f>\frac{1}{2}\mathrm{d}mmp/\mathrm{d}f$.[9]

Although in our model the uninformed traders are hurt by preferencing, individual uninformed traders cannot decrease their costs by choosing where to trade.[10] As a result, they are indifferent between trading in the competitive portion or the preferenced portion of the market. The question remains whether some form of competition outside the model might serve to reduce their costs. For example, competition to provide brokerage to the uninformed could lead to lower costs or increased services. We examine spreads in preferenced market centers in the empirical analysis in Section 4. The data do not allow us to measure other costs or benefits such as commissions or rebates. Note, however, that provision of brokerage services is a declining marginal cost business, making it difficult for new entrants to compete with established brokers. Thus, it is not clear whether this type of competition would necessarily eliminate the broker/market maker profits from preferencing.

Regulators have reason to be concerned with losses borne by either informed or uninformed traders. Informed traders make prices more accurate, and uninformed traders supply investment capital. Lower expected profits for informed traders may reduce their incentive to search for costly information, thereby reducing market efficiency. Higher trading costs for uninformed traders may reduce their propensity to invest. Our model does not explicitly capture either of these effects because we do not model the uninformed investors' initial decisions to invest, the informed investor's decisions to acquire costly information, or firms' use of market prices in their real asset investment decisions.

3. Data and methods

Guided by the results in our theoretical model, we compare market quality for NASDAQ and NYSE-listed stocks using monthly execution quality reports published pursuant to SEC Rule 11Ac1-5. Our sample period covers January through June 2004. To construct our sample of stocks, we start with all U.S. common stocks in the University of Chicago's Center for Research in Security Prices (CRSP) database that are listed either on the NYSE or on NASDAQ's national market system. (NASDAQ small-cap stocks are not covered by Rule 11Ac1-5.) We include only stocks that meet the NYSE's continued listing requirements as of the beginning of the sample period.[11]

One potential difference between NASDAQ and NYSE-listed stocks is that more of the NASDAQ stocks may be recent IPOs, and it may be more difficult to make a market in these stocks. To ensure that the stocks have had reasonable "seasoning" prior to inclusion in our sample, we add several screens. We eliminate all stocks that are not in Standard and Poor's Compustat database as of two years prior. We also eliminate stocks that have incomplete daily return information or had not been classified as a common stock for the preceding 24 months. We further exclude stocks that split or change their ticker symbol or CUSIP number during the sample period. Finally, we require each stock to have 11Ac1-5 execution quality statistics for marketable orders in each order size category for each

[9]See "Result 3" in Appendix A.

[10]This statement assumes there are many uninformed traders so one individual trader moving his or her order between markets will not materially affect the overall preferencing fraction.

[11]See www.nyse.com/listed for a description of the NYSE listing standards.

month in the sample. The resulting sample consists of 1027 NYSE-listed stocks and 951 NASDAQ stocks.

In comparing NYSE-listed stocks to NASDAQ stocks, we first separate the stocks into the following four market capitalization categories: below \$200 million, between \$200 million and \$1 billion, between \$1 billion and \$10 billion, and above \$10 billion.[12] Within these market capitalization categories we control for the following attributes:

- Market capitalization as of December 31, 2003.
- Average share price (computed using daily bid-ask midpoints as of 11:00 a.m.).
- Total executed share volume across all 11Ac1-5 reporting market centers, for market and marketable limit orders.[13]
- Standard deviation of daily percentage changes in the bid-ask midpoint. (Bid-ask midpoints are measured as of 11:00 am.)
- Whether the stock passes the NYSE's *initial* listing requirements as of December 31, 2003. (Recall that all stocks in our sample pass the NYSE's *continued* listing requirements.)

Table 1 reports the characteristics of the firms in our sample for each market capitalization category.

3.1. Execution quality statistics

Execution quality statistics come from the monthly electronic reports required of all market centers that trade national market system (NMS) securities under SEC Rule 11Ac1-5.[14] The term *market center* refers to specialists on the NYSE or regional exchanges, NASDAQ dealers, electronic communications networks (ECNs), or alternative trading systems. Each market center's report includes statistics for every NMS stock it trades. For example, execution quality statistics for NYSE-listed stocks appear not only in NYSE specialists' reports, but also in the reports of third-market dealers, regional exchanges, and other trading systems.

Each market center computes its execution quality statistics for all covered orders—orders for fewer than 10,000 shares that are submitted during regular trading hours and request no special handling.[15] All statistics reflect the shares that were executed by the receiving market center, together with shares that were routed out by the receiving market center and executed at another venue.

The reports provide statistics for each stock by order type and order size. There are five order-type categories: market orders, marketable limit orders, inside-the-quote limit orders, at-the-quote limit orders, and near-the-quote limit orders. We examine only marketable orders (i.e., market orders and marketable limit orders) since these are most

[12]These size categories are the same as those used in SEC (2001).

[13]The volume calculations exclude orders that are routed out by the receiving market center. Such orders appear in the reports of both the receiving market center and the executing market center, which would result in double-counting of this order flow were these orders included. See Section 3.2 for more discussion.

[14]National market system securities include exchange-listed equities and NASDAQ National Market tier securities. (See Securities Exchange Act Rule 11Aa2-1.)

[15]Securities Exchange Act Release No. 43590 (November 17, 2000), 65 FR 75414 ("Adopting Release") provides details.

Table 1
Characteristics of sample firms

We select all NYSE-listed and NASDAQ common stocks for which we have 11Ac1-5 data for marketable orders in all four order size categories for each of the months from January 2004 through June 2004. We further require that all firms pass the NYSE's *continued* listing requirement as of the start of the sample, have no symbol or CUSIP number changes or stock splits from January 2004 through June 2004, have complete daily return information for the preceding 24 months, were identified as a common stock for the past 24 months, and were in Compustat as of two years prior to the start of the sample. Market Capitalization and the dummy variable indicating whether the firm passes the NYSE initial listing requirement are measured as of December 31, 2003. Means reflect equally weighted averages across the six months in the sample, medians are averages of the monthly cross-sectional medians, and the standard deviations are averages of the monthly cross-sectional standard deviations.

	>\$10 Bil.		\$1–\$10 Bil.		\$0.2–\$1 Bil.		<\$0.2 Bil.	
	NASDAQ	NYSE	NASDAQ	NYSE	NASDAQ	NYSE	NASDAQ	NYSE
Number of firms	28	181	214	500	426	285	283	61
Fraction that pass NYSE init. listing	0.964	0.950	0.706	0.928	0.369	0.642	0.102	0.197
Market capitalization (millions)								
Mean	46,158	36,487	2,730	3,442	457	568	114	112
Median	17,746	18,893	1,938	2,666	404	559	112	112
Std. dev.	66,007	47,697	1,998	2,269	211	221	45	47
Share price								
Mean	40.26	46.93	30.88	34.73	19.26	20.83	9.40	8.07
Median	38.63	43.49	29.50	32.35	16.82	18.73	7.67	6.48
Std. dev.	17.39	21.72	17.24	16.89	11.44	11.59	6.50	4.93
Bid-ask midpoint volatility								
Mean	0.0185	0.0141	0.0231	0.0168	0.0285	0.0217	0.0354	0.0294
Median	0.0176	0.0128	0.0226	0.0156	0.0279	0.0200	0.0328	0.0274
Std. dev.	0.0058	0.0053	0.0085	0.0060	0.0100	0.0079	0.0282	0.0117
Marketable order shares (millions)								
Mean	183.0	47.6	27.2	11.1	5.0	3.7	2.6	2.1
Median	79.1	31.5	10.1	8.0	2.4	2.3	1.2	0.8
Std. dev.	239.2	54.3	55.6	12.1	7.5	5.0	5.4	4.0

likely to be preferenced. Within each order-type category, the reports present statistics for four order-size categories: 100–499 shares, 500–1999 shares, 2000–4999 shares, and 5000–10,000 shares.

For each order-size category, the reports present the number of orders and the total number of shares, along with several execution quality measures including effective and realized spreads and time to execution. The effective spread for a buy order is calculated by subtracting the midpoint of the national best bid and offer (NBBO) quoted spread at the time the order arrived at the market center from the transaction price, and then multiplying this difference by two. Effective spreads for sell orders are computed analogously. The effective spread is an estimate of the execution cost actually paid by traders. Realized spreads, which are computed using the quote midpoint five minutes after the trade is executed rather than the current midpoint as the benchmark, capture liquidity suppliers' revenue net of losses to better-informed traders. We compute the preferencing measure (*pm*) by taking the ratio of realized to effective spreads. For completeness, we also examine quoted spreads, time to execution, and execution rates.[16]

We obtain 11Ac1-5 reports from two independent companies—Transaction Auditing Group (www.tagaudit.com) and Market Systems, Inc. (www.marketsystems.com). These firms specialize in best execution analysis and have gathered execution quality reports and posted them online. They have no obligation to maintain reports for all market centers, but it is to their advantage to provide coverage that is as comprehensive as possible.

After applying standard filters to our data to eliminate likely errors, we compute share-weighted six-month averages across all market centers for each execution quality measure for all marketable orders in each order-size category for each stock.[17] Although the execution quality statistics reported by each market center also reflect orders that are routed out, we omit these shares when computing the weights, so as not to double-count orders reported by both the receiving market center and the executing market center. Our aggregate execution quality measures will accurately reflect true execution quality as long as orders that are routed out receive similar execution to those that are executed by the receiving market center.[18] The fact that less than 5% of orders in NYSE-listed stocks are routed out suggests that any differences in execution quality for executed and routed-out shares would have a limited impact on overall averages for these stocks. The fraction is larger (18%) for NASDAQ stocks, but routed-out orders are concentrated at market centers with built-in order-routing facilities. These services receive orders and automatically route them to the market center with the best price, so there is no reason to believe that execution quality is necessarily poorer for routed-out shares.

Table 2 categorizes marketable order executions from June 2004 into market order executions and marketable limit order executions for several of the large market centers in both NASDAQ and NYSE-listed stocks. We have combined order-size categories to save space. Table 2 shows that the NYSE has a dominant position in trading NYSE-listed stocks, but no single market center has a dominant position in trading NASDAQ stocks. The largest market center for NASDAQ stocks is INET (the new entity created from the merger of the Island and Instinet ECN's.), with 38% of marketable orders. The total

[16]Although the NBBO spread is not included in the reports, we can calculate it using the following relationship: Quoted spread = effective spread + average price improvement*2. Average price improvement is computed using the 11Ac1-5 reports and reflects orders executed both inside the spread (positive price improvement) and outside the quotes (negative price improvement). The time to execution is computed as the share-weighted average of the time to execution for three separate categories of orders—shares executed at the quote, shares executed inside the quoted spread, and shares executed outside the spread. Execution rates are computed as the total shares executed across all market centers divided by the total shares submitted.

[17]We delete any observations where the absolute effective spread exceeds 50% of the stock price, or the quoted spread exceeds $3. We also eliminate observations with realized spreads greater than the maximum daily price change for the stock for the current month. Finally, we discard observations in which the time to execution is negative or exceeds 6.5 hours. (We assume that execution times of –1 s are the result of clock synchronization problems. Accordingly, we set these values to zero before applying the filter.) Data from Island ECN (ISLD) are clearly erroneous in March 2004 and missing in April 2004. These months are excluded when computing six-month share-weighted average execution quality measures. For marketable orders, we compute center-executed volume by subtracting routed out shares from the sum of shares executed inside, at, and outside the quotes. In most cases, this is equivalent to simply using the number of shares executed at the market center as stated in the 11Ac1-5 reports. (These two approaches would always yield identical results in the absence of reporting inconsistencies.)

[18]We also compute measures weighting by the total shares for each market center (including those routed out and executed at another market center) and the results (not shown) are similar.

volume figures reveal that marketable limit orders are an important component of marketable orders, constituting over 60% of all marketable order executions for NYSE-listed stocks and more than 85% of all marketable order executions for NASDAQ stocks.

The fact that NYSE-listed and NASDAQ stocks have different proportions of market and marketable limit orders suggests that comparisons between the two markets that focus

Table 2
11Ac1-5 executions of marketable orders

Market center executions of market and marketable limit orders, as a percent of total for the NYSE-listed and NASDAQ stocks in our sample for June 2004.

Market center	100–1999 Share orders			2000+ Share orders			All
	Market	Mktable limit	Total mktable	Market	Mktable limit	Total mktable	mktable orders
Panel A: NYSE-listed stocks							
NYSE	78.1%	93.5%	87.6%	74.4%	83.6%	80.0%	85.3%
Knight Capital Markets	4.5	1.5	2.6	5.0	5.2	5.1	3.4
Bernard Madoff	4.3	0.2	1.8	3.9	0.8	2.0	1.9
Midwest Ex.	2.5	0.6	1.3	3.0	3.2	3.1	1.9
Cincinnati	3.1	0.2	1.3	5.1	1.5	2.9	1.8
Boston	2.9	0.3	1.3	3.5	1.1	2.0	1.5
Schwab Capital Markets	2.0	0.2	0.9	2.5	0.6	1.4	1.0
INET	0.0	1.7	1.0	0.0	1.2	0.7	0.9
Pacific Ex./Archipelago	0.0	1.4	0.9	0.0	1.0	0.6	0.8
Automated Trading Desk	0.9	0.1	0.4	1.2	0.7	0.9	0.6
TD Waterhouse	0.5	0.0	0.2	0.7	0.4	0.5	0.3
Philadelphia Ex.	0.2	0.1	0.2	0.3	0.5	0.4	0.2
Citigroup Global Markets	0.5	0.0	0.2	0.2	0.0	0.1	0.2
Moors & Cabot	0.1	0.0	0.1	0.2	0.1	0.1	0.1
Brut	0.0	0.1	0.1	0.0	0.1	0.0	0.1
Other NYSE-listed	0.3	0.0	0.1	0.1	0.1	0.1	0.1
Total NYSE-listed	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Volume (mil)	3687	6042	9730	1682	2638	4319	14049
Panel B: NASDAQ stocks							
INET	0.0%	39.1%	33.6%	0.0%	49.8%	43.6%	38.2%
Pacific Ex./Archipelago	11.8	24.3	22.5	10.0	15.6	14.9	19.0
NASDAQ SuperMontage	1.2	21.3	18.5	0.7	16.9	14.9	16.9
Brut	3.8	7.8	7.3	1.9	4.4	4.1	5.8
Knight Equity Markets	18.3	1.3	3.7	14.2	3.5	4.9	4.2
Schwab Capital Markets	14.6	1.0	2.9	23.2	2.3	5.0	3.8
B-Trade Services	0.0	2.7	2.3	0.1	1.8	1.6	2.0
Automated Trading Desk	5.9	0.4	1.2	5.6	0.6	1.2	1.2
GVR Company	3.6	0.2	0.7	5.2	0.7	1.3	1.0
National Financial Services	7.1	0.0	1.0	6.6	0.1	0.9	1.0
Attain	0.0	0.8	0.6	0.0	1.4	1.2	0.9
Morgan Stanley	3.8	0.1	0.6	3.9	0.3	0.8	0.7
Citigroup Global Markets	3.5	0.1	0.5	5.0	0.2	0.8	0.7
Bernard Madoff	4.8	0.1	0.7	2.2	0.1	0.4	0.6
Pershing Trading Co.	2.4	0.1	0.4	3.1	0.2	0.6	0.5
Other NASDAQ	19.1	0.8	3.3	18.3	2.0	4.0	3.6
Total NASDAQ	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Volume (mil)	1,004	6,215	7,220	770	5,367	6,137	13,356

on either order type in isolation may be misleading. Both market and marketable limit orders allow investors to trade with high probability of execution. If investors' choices between these two types of orders depend on market conditions and these conditional decisions differ across markets, then a comparison of either type of order in isolation will mean comparing the markets based on different market conditions. The distinction between market and marketable limit orders is further blurred because brokers sometimes convert market orders into marketable limit orders to route them to an ECN. For example, the fraction of marketable limit orders is likely inflated because INET does not accept market orders. The 11Ac1-5 reports reflect the order type as it was received at the market center, not the order type originally submitted by the customer.

Our focus on the combination of all marketable orders differs from Boehmer (2005) who uses just market orders. He points out that a fairly substantial fraction of marketable limit orders do not execute. If the reason these orders fail to execute is that the market is "moving away," then one would want to make some sort of adjustment for the opportunity cost associated with these failed executions. On the other hand, some of the non-executions occur for other reasons. For example, ECNs sometimes contain aggressively priced orders that are not displayed in the quote. As a result, a trader who wants to buy immediately may submit a limit order to an ECN that is not matching the NBBO ask price to see if it crosses with one of these hidden orders. (Rule 11Ac1-5 categorizes this order as marketable if its limit price matches the NBBO ask quote at the time of submission.) If the trader's order does not execute almost immediately, the trader may cancel it and send a market order to a different market center. Hasbrouck and Saar (2002) call these short-term marketable limit orders sent to ECNs "fleeting" orders. By including the combined realized spreads from executed market and marketable limit orders, we capture the total average cost of traders who follow this two-step strategy, but if we exclude the marketable limit order executions, then we miss the cases where there was a hidden order in the ECN.

In subsequent tests, we focus on order sizes between 100 and 499 shares and between 500 and 1,999 shares because these two groups are most likely to be covered by preferencing agreements. Moreover, the 11Ac1-5 reports exclude orders that request special handling, so the statistics for the larger order sizes may be less reliable, especially for cross-exchange comparisons. These special orders make up a much greater fraction of large orders than of small orders (e.g., the "not held" orders often used by large institutional traders), so a substantial portion of the orders in the larger size categories may be excluded from the data, which limits our ability to draw inferences from the results. Furthermore, possible definitional differences in special handling between NASDAQ and the NYSE may result in exclusion of different types of orders from the 11Ac1-5 reports for each group of stocks. Finally, to the extent that traders with larger orders use different criteria across the two markets to determine whether to request special handling, the resulting 11Ac1-5 samples would not be directly comparable.

3.2. *Regression methods*

To fairly compare execution quality for NASDAQ and NYSE-listed stocks, we need to control for differences in characteristics known to affect liquidity and adverse selection. For each firm/order-size category, we regress each execution quality measure on an exchange dummy (that equals 1 for NASDAQ stocks) and control variables including the natural logarithm of market capitalization as of 31 December 2003 (to control for variation within the firm-size category); the inverse of the average share price; the natural logarithm of share volume; bid-ask midpoint volatility over the six-month sample period; and a dummy variable that assumes a value of one if the stock satisfies the *initial* listing requirements on the NYSE. (All firms pass the continued listing requirements.) Coefficients on the exchange dummy variable measure differences in execution quality across exchanges after controlling for other differences in the stocks.

4. Results

Table 3 presents coefficient estimates and p-values for the exchange dummy variable for each order size and firm size category. To provide an indication of the typical magnitudes of the various statistics, the table also reports predicted values. For both NYSE and NASDAQ stocks, the predicted values are computed using the median values of the continuous control variables for NASDAQ stocks, as reported in Table 1. The initial listing requirement dummy variable is set to the fraction of NASDAQ stocks in the sample that pass the NYSE initial listing requirements, and the exchange dummy variable is set to 1 for NASDAQ stocks and 0 for NYSE stocks Thus, the estimated exchange dummy coefficient is equal to the difference between NASDAQ and NYSE predicted values.

Quoted spreads are significantly wider for NASDAQ stocks than for NYSE-listed stocks for all but the largest firms in the sample. This is consistent with Bessembinder's (2003a) results using transaction data in the post-decimal period. Effective spreads for NASDAQ stocks are wider than those for NYSE-listed stocks in each firm size and order size category. In addition, we find that effective spreads exceed quoted spreads for NASDAQ stocks in all firm size categories, which is consistent with the results Bessembinder (2003a) reports for the largest NASDAQ stocks.

Realized spreads, which our model shows to be a better measure of market quality in the presence of preferencing agreements, are also uniformly higher for NASDAQ stocks. Specifically, realized spreads for NASDAQ stocks exceed those for comparable NYSE stocks by as much as $0.025 per share on average. Using 11Ac1-5 data from November 2001 through December 2003, Boehmer (2005) also finds that NASDAQ realized spreads are wider for market orders. (As discussed above, he does not consider marketable limit orders.)

The difference between the effective spread and the realized spread is a measure of the amount of information contained in the order flow. Comparison of effective and realized spreads in Table 3 indicates that adverse selection is generally higher for NYSE-listed than for NASDAQ stocks. This is consistent with work by Affleck-Graves et al. (1994) and Lin et al. (1998).

Realized spreads for NYSE-listed stocks are negative for many firm/order-size categories. This result is driven by negative realized spreads in the primary market (as shown in Table 4 and discussed below).[19] Negative realized spreads are inconsistent with our model, because we assume that both informed and uninformed traders submit marketable orders, and liquidity suppliers are either competitive or preferenced market

[19]Consistent with the evidence in Lipson (2004), the negative realized spreads for NYSE stocks stem from marketable limit orders. Average realized spreads for NYSE stocks in the three largest firm-size categories range from −0.008 to −0.005 for marketable limit orders, but are uniformly positive for market orders.

Table 3
Execution quality measures for 100–499 and 500–1999 share marketable orders

For each market capitalization category and each of the two smaller order size categories, we regress each stock's six-month share-weighted average for each measure on the following control variables: an intercept, a dummy variable that equals one if the stock is a NASDAQ stock, a dummy variable that equals one if the stock passes the NYSE *initial* listing requirements, the log of market capitalization, the inverse of share price, the log of average monthly executed 11Ac1-5 marketable order volume, and average monthly bid/ask midpoint volatility. The table reports predicted values from this regression using the fraction of NASDAQ stocks that pass the initial listing requirement and the median values of the other control variables for NASDAQ stocks reported in Table 1. The difference between the NASDAQ and NYSE predicted values is the estimated coefficient on the dummy variable, and the *p*-values for this estimate are reported in parentheses.

	Market capitalization categories							
	>$10 billion		$1–$10 billion		$0.2–$1 billion		<$0.2 billion	
	100–499	500–1999	100–499	500–1999	100–499	500–1999	100–499	500–1999
Quoted spreads								
NASDAQ	0.012	0.013	0.026	0.029	0.046	0.050	0.057	0.057
NYSE	0.014	0.014	0.021	0.022	0.036	0.038	0.044	0.047
diff.	−0.002	−0.001	0.004	0.007	0.011	0.012	0.013	0.010
	(0.007)	(0.459)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.006)
Effective spreads								
NASDAQ	0.017	0.019	0.029	0.035	0.047	0.058	0.056	0.063
NYSE	0.012	0.013	0.018	0.025	0.029	0.046	0.035	0.051
diff.	0.005	0.007	0.011	0.010	0.018	0.012	0.021	0.012
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)
Realized spreads								
NASDAQ	0.006	0.005	0.011	0.007	0.024	0.016	0.035	0.024
NYSE	−0.001	−0.003	−0.004	−0.005	−0.001	−0.004	0.016	0.014
diff.	0.006	0.008	0.015	0.012	0.025	0.020	0.018	0.010
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Preferencing measure								
NASDAQ	0.378	0.294	0.338	0.206	0.465	0.287	0.620	0.405
NYSE	0.006	−0.142	−0.204	−0.223	−0.076	−0.115	0.447	0.282
diff.	0.372	0.436	0.543	0.430	0.541	0.401	0.173	0.123
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Execution time (*s*)								
NASDAQ	1.5	3.2	1.2	5.7	2.5	18.3	4.9	32.7
NYSE	8.4	10.3	11.8	17.5	15.7	41.3	19.1	60.7
diff.	−6.9	−7.1	−10.6	−11.8	−13.2	−23.0	−14.2	−27.9
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Execution rate								
NASDAQ	0.669	0.587	0.762	0.599	0.810	0.632	0.813	0.725
NYSE	0.757	0.801	0.782	0.813	0.813	0.840	0.847	0.889
diff.	−0.088	−0.214	−0.020	−0.214	−0.003	−0.208	−0.034	−0.165
	(0.000)	(0.000)	(0.000)	(0.000)	(0.454)	(0.000)	(0.002)	(0.000)

makers. In our model, uninformed traders pay positive realized spreads, informed traders have negative realized spreads, and the average realized spread in the competitive portion of the market is zero because market makers set quotes to just break even.

Table 4
Effective and realized spreads for non-primary market centers compared to the primary market center for 100–499 and 500–1999 share orders

We calculate share-weighted average effective and realized spreads for each market center across the two smallest order-size categories for January–June 2004. We define the primary market center as the NYSE for NYSE-listed stocks and SuperMontage for NASDAQ stocks. For each primary market center, the table reports averages for all stocks that had at least 10,000 shares executed at the primary market center over the sample period. For each non-primary market center, the table reports averages of stock-by-stock differences in realized and effective spread between the primary and non-primary market center across only those stocks in the market capitalization category that had at least 10,000 shares executed at both the primary and the non-primary market center. A positive value means that the non-primary market center is higher. Statistical significance is based on t-tests using the across-stock standard deviation of these differences, and entries that are significant at the 1% (5%) level in a two tailed test are identified by *(+).

Market center	Market capitalization categories							
	>$10 billion		$1–$10 billion		$0.2–$1 billion		<$0.2 billion	
	Effect. spread	Real. spread	Effect. spread	Real. spread	Effect. spread	Real. spread	Effect. spread	Real. spread
NYSE	0.017	−0.002	0.021	−0.004	0.030	−0.004	0.043	0.013
SuperMontage	0.015	0.001	0.031	0.005	0.053	0.017	0.059	0.023
Differences relative to the NYSE								
Knight	0.001*	0.007*	0.007*	0.011*	0.013*	0.010*	0.005*	0.003*
Madoff	0.001*	0.015*	0.004*	0.015*	0.005*	0.014*	0.002*	0.005+
Midwest	0.005*	0.013*	0.010*	0.013*	0.013*	0.010*	0.012*	0.008*
Cincinnati	−0.004*	0.010*	−0.002*	0.014*	−0.001*	0.010*	0.005*	0.012
Boston	0.005*	0.016*	0.008*	0.016*	0.010*	0.012*	0.006*	0.010*
Schwab	0.002*	0.018*	0.007*	0.024*	0.013*	0.027*	0.009*	0.083
INET	0.008*	0.012*	0.015*	0.030*	0.015*	0.071*	0.007*	0.014
Pacific	0.004*	−0.001	0.004*	−0.003*	−0.001*	−0.003*	−0.001*	−0.007+
Auto.Trading	0.002*	0.010*	0.004*	0.014*	0.002*	0.015*	0.003*	0.009*
TD Waterhouse	0.002*	0.014*	0.005*	0.012*	0.005*	0.009*		
Philadelphia	0.003*	0.002	0.009*	0.003	0.009*	−0.002	0.005*	−0.020
Citigroup	−0.005*	0.014*	−0.003*	0.016*	−0.003*	0.015*	−0.002*	0.008
Moors & Cabot	0.008*	0.019*	0.017*	0.021*	0.012*	0.023		
Differences relative to SuperMontage								
INET	0.011	0.002	−0.002*	0.001	−0.005*	0.000	−0.003*	0.001
Pacific	0.006*	0.010*	0.003*	0.009*	−0.001*	0.008*	−0.004*	0.003*
Brut	−0.001*	0.002*	−0.003*	0.002*	−0.005*	0.002+	−0.001*	0.003*
Knight	0.001*	0.012*	0.002*	0.016*	0.002*	0.013*	−0.002*	0.011*
Schwab	−0.000	0.010*	0.003*	0.014*	0.007*	0.015*	0.005*	0.015*
B-Trade	0.000	−0.001	−0.006*	−0.003*	−0.010*	−0.008*	−0.011*	−0.010*
Auto.Trading	0.001*	0.005*	0.003*	0.004*	0.004*	0.004	0.001*	0.006
GVR Co.	0.001+	0.006*	0.002*	0.005+	0.001*	0.011*	−0.005*	0.010*
Fidelity	0.000	0.010*	0.001*	0.013*	0.001*	0.012*	−0.001*	0.009*
Attain	−0.004*	0.003	−0.005*	−0.001	−0.010*	−0.003	−0.005*	−0.006
Morgan Stanley	0.001+	0.012*	0.005*	0.012*	0.015*	0.018*	0.010*	0.011*
Citigroup	0.000	0.012*	0.002*	0.010*	0.001*	0.010*	−0.002*	0.003
Madoff	0.001	0.012*	0.001*	0.010*	−0.001*	0.009*	−0.002*	0.007+
Pershing	0.000	0.010*	0.000*	0.009*	−0.004*	0.005*	−0.003*	0.002

In reality, both informed and uninformed traders may use other strategies, particularly in the competitive portion of the market. For example, uninformed traders may submit limit orders and later use marketable orders if their initial limit orders fail to execute. Such traders may be willing to use aggressively priced limit orders to increase the likelihood of execution, which would lead to negative average realized spreads for marketable orders. The 11Ac1-5 reports also include realized spreads for non-marketable limit orders. As with marketable orders, these realized spreads are computed by comparing the execution price to the quote midpoint five minutes after execution. We find that average realized spreads are *positive* for the executed *non-marketable* limit orders in NYSE-listed stocks included in the 11Ac1-5 reports, which suggests that these particular liquidity suppliers are indeed willing to bear some cost to trade.

Expanding our model to incorporate "aggressive" liquidity suppliers would require assumptions about the timing and motivation for their trades and would substantially complicate the trading game. Roughly speaking, we would expect their presence to reduce both the effective and realized spreads paid by the marketable orders. Thus, their presence would change the interpretation of the *level* of our preferencing measure, but our primary focus is on the *difference* between the preferencing measures across markets. If the proportional impacts of the aggressive liquidity suppliers on the effective spreads (the denominators in the measures) are relatively small, then their primary effect on the measures will be due to changes in realized spreads (the numerators in our measures). If it is also true that the impacts of aggressive liquidity suppliers on realized spreads are similar across the two markets, then we can still interpret the difference between the preferencing measures as a proxy for the difference between the degrees of preferencing across the markets.

The next portion of Table 3 compares our new preferencing measure (pm) across the two markets. The results suggest that the poorer execution quality for small orders in NASDAQ stocks is indeed related to preferencing. Pm is uniformly higher for NASDAQ stocks. The magnitude of the difference between the preferencing measure for NYSE-listed and NASDAQ stocks suggests that dealers in NASDAQ stocks capture additional profits of up to 55% of the effective spread on average.

Chung et al. (2004) examine preferencing in NASDAQ stocks and find that between 62 and 76 percent of volume was preferenced on average in June 2001, with preferencing fractions by stock assuming the full range from 0 to 1. Although the values for pm in Table 3 are somewhat lower, this may reflect the presence of "aggressive" liquidity suppliers, as discussed above. Also, recall that our pm measure is an estimate of $f/(2-f)$. An alternative preferencing measure discussed in Section 2.3 is given by

$$pm_a = f = \frac{2pm}{(1+pm)} = \frac{2\left(\overline{rs}^p/\overline{es}^p\right)}{\left(1+\left(\overline{rs}^p/\overline{es}^p\right)\right)}.$$

Although this measure is difficult to estimate reliably on a stock-by-stock basis, we can mitigate these problems by using the predicted values for each firm/order-size group. The resulting values range from 0.33 to 0.77, which are closer to the values reported by Chung et al. (2004).

Execution times for orders in NASDAQ stocks are lower than for orders in NYSE-listed stocks, also consistent with Boehmer's (2005) results.[20] Speed is presumably an important

[20]Boehmer (2005) suggests that inclusion of marketable limit orders makes it difficult to interpret execution time comparisons across markets because some orders do not execute and others execute with a substantial delay. In

aspect of execution quality to traders with short investment horizons and to investors who base trading decisions on time-sensitive information. Consequently, shorter execution times may suggest better market quality for NASDAQ stocks. At the same time, the results demonstrate that speed is clearly not synonymous with guaranteed immediate execution, since execution rates are uniformly lower for NASDAQ stocks than for NYSE-listed stocks. Moreover, shorter execution times may reduce the ability of orders to interact, thereby raising execution costs.

We check the robustness of our results by running random effects panel data regressions using monthly share-weighted averages of the execution quality statistics.[21] Although the panel data regression specification is somewhat less restrictive, the monthly estimates are noisier than the six-month averages. The results for the panel data regressions (not shown) are similar to those in Table 3.

As an additional robustness check, we run our ordinary least squares (OLS) regressions using 11Ac1-5 data from January through June 2002 (results not shown). Between 2002 and 2004, spreads have narrowed and execution times have declined for both NYSE-listed and NASDAQ stocks. In spite of this decline, the differences in spreads across exchanges in 2002 are similar to the differences reported in Table 3. Although Table 3 shows that NASDAQ execution times are lower for all firm/order-size categories in 2004, in 2002, the NYSE execution times were lower for small orders in all but the largest firm-size category. The 2002 results for the preferencing measure and execution rates are similar to the 2004 results presented in Table 3, both in terms of levels and differences between the NYSE and NASDAQ.

Finally, we address potential small-sample bias in our execution quality measures. In Section 2.1, we show that the share-weighted average effective spreads reported in the 11Ac1-5 data converge to the ratio of the expected aggregate effective spread to the expected gross volume in our model as the number of repetitions of the model (T) approaches infinity. The fact that the denominator of this ratio is estimated with error for finite T implies that the ratio will be biased upward (by Jensen's inequality). Aggregating across several months should reduce the estimation error, mitigating this problem. As an additional robustness check, we also apply an explicit bias adjustment to all estimated ratios, and obtain results similar to those presented in Table 3. (See Appendix B for details of the bias adjustment.)

The results of the above analysis demonstrate that execution quality remains poorer for NASDAQ than for NYSE-listed stocks in the post-decimal trading environment and suggest that the difference is driven by more preferencing-induced fragmentation for NASDAQ stocks.

4.1. Execution quality by market center

Our model assumes that preferenced market centers match the effective spreads paid by *uninformed* traders at other market centers, which means that the average effective spreads

(*footnote continued*)
spite of these concerns, our execution time results for all marketable orders are similar to Boehmer's results using only market orders.

[21]For the panel data regressions, we eliminate any stock-month observations with fewer than 20 orders to increase the precision of the execution quality statistics.

in the preferenced market centers should be lower than the overall average for the competitive market centers, which include both informed and uninformed orders. Our model also predicts that preferenced market centers will have higher realized spreads than non-preferenced market centers. To investigate these two predictions, we compare the effective and realized spreads for each market center to those of the primary market. We define the primary market as the NYSE for NYSE-listed stocks and NASDAQ's SuperMontage for NASDAQ stocks. For each non-primary market center, we recompute benchmark spreads for the primary market using only the stocks that are traded by that particular market center. This ensures that differences in spreads are not driven by the characteristics of the subset of stocks traded by the non-primary market center.

Table 4 reports average realized and effective spreads for the primary markets, as well as differences in spreads between the primary and non-primary market centers. The non-primary market centers are shown in the same order as in Table 2. Differences are computed relative to the primary market, so a positive value means that the spread on the non-primary market is higher. To save space, statistics have been combined for the two smaller order-size categories.

The first rows of Table 4 show that average realized spreads on the NYSE and NASDAQ's SuperMontage tend to be close to zero or even slightly negative, particularly for the larger firm-size categories. The remainder of the table examines non-primary market centers. Broadly speaking there are three types of non-primary market centers represented. Firms like Fidelity, which has most of the Boston Exchange volume in listed stocks, and Charles Schwab are internalizers. Although they also accept orders from other sources, the bulk of their 11Ac1-5 orders probably come from their own brokerage operations. Knight and Madoff do not have their own brokerage arms, but have preferencing arrangements with various brokers that include payment for order flow. The third category of market centers are the Electronic Communications Networks (ECN's) like INET and the Pacific Exchange (where Archipelago provides the trading platform). The ECN market centers themselves do not engage in preferencing.

Table 4 shows that average realized spreads on most of the non-primary market centers are significantly wider than those in the primary market center, which is not surprising because much of this order flow is covered by some type of preferencing arrangement. The fact that realized spreads are wider on some ECN's is somewhat surprising, and may indicate that profitable trading opportunities exist on ECN's.

Although Table 4 shows that realized spreads at the alternate market centers are generally wider than on the primary market center, we also find in results not shown that the realized spreads are lower than the effective spreads charged by the same market center. The fact that realized spreads are not equal to effective spreads implies that the preferenced dealers are not entirely successful in attracting purely uninformed order flow. For example, informed traders may break up their orders to satisfy the screens in place at the preferenced market centers. This is consistent with the Chung et al. (2004) finding that the price impact of internalized orders on NASDAQ, while lower than that of non-preferenced orders, is non-zero. Differences between realized and effective spreads may also reflect the fact that few market centers obtain 100% of their order flow through preferencing agreements. Bessembinder's (2003b) result that dealers in off-board markets for NYSE-listed stocks earn much lower realized spreads when they post quotes that match the NBBO (which they do with some frequency) suggests that order flow obtained outside preferencing agreements does in fact contain more information.

Table 4 shows that the non-primary market center effective spreads are generally wider than those in the primary market center. These results are consistent with work by Lipson (2004), who uses 11Ac1-5 data from July 2001 through June 2002 to compare the trading of NYSE-listed securities on the primary exchange and in other markets. The wider effective spreads in the non-primary market centers are inconsistent with the more favorable terms assumed in our model, but may also be related to the fact that some of the orders sent to these market centers are not covered by preferencing arrangements.

5. Conclusion

We present a model that considers the impact of preferencing on two common measures of execution quality—effective spreads and realized spreads — and demonstrate that both uninformed and informed investors are hurt when preferenced market makers match the effective spreads given to uninformed orders in the primary market. The model also shows that effective spreads underestimate the true costs of preferencing, and that realized spreads are a better measure of market quality in the presence of different degrees of preferencing. On the other hand, we show that realized and effective spreads are equally sensitive to differences in information asymmetry, so even comparisons of realized spreads can be problematic when the degree of adverse selection differs.

We propose a new preferencing measure that reflects the fraction of uninformed order flow that has been successfully captured by preferenced market makers. Our measure can be used to quantify the profits earned by preferenced dealers, and controls for the degree of information asymmetry in the market, which makes it ideal for cross-exchange comparisons of market quality.

Our empirical tests compare the preferencing measure and effective and realized spreads (along with quoted spreads, execution speed, and execution rates) for marketable orders in NYSE-listed and NASDAQ stocks from January through June 2004 using the monthly execution quality reports required by SEC Rule 11Ac1-5. Regression analysis indicates that execution quality for orders of less than 2000 shares is poorer for NASDAQ stocks than for NYSE-listed stocks in the post-decimal trading environment. Realized and effective spreads are narrower for NYSE-listed stocks in all firm-size categories, and quoted spreads are narrower for all but the largest firms in the sample. The higher preferencing measure for NASDAQ stocks confirms that the poorer execution quality is related to a higher level of preferencing, and indicates that dealers in NASDAQ stocks capture additional profits of up to 55% of the effective spread. Examination of results by market center demonstrates that differences in realized spreads are more pronounced than differences in effective spreads, and preferenced market centers have higher realized spreads than non-preferenced market centers.

The results demonstrate a direct link between poorer execution quality and a higher degree of preferencing-induced fragmentation. Preferenced dealers can and do earn substantial profits at the expense of investors by successfully diverting uninformed order flow from the primary market. This means preferencing is a potentially harmful feature of U.S equity markets, and that regulators should consider ways to control or discourage it.

Appendix A. Proofs of model results

Result 1. Net order flow $\tilde{B}-\tilde{S}=\tilde{x}+\tilde{u}$.

Proof.

Total gross order flow

$$\tilde{G}=\sum_{i=1}^{N_u}|\widetilde{u}_i|+|\widetilde{x}|.$$

Gross buy orders

$$\tilde{B}=\sum_{i=1}^{N_u}\max[\widetilde{u}_i,0]+\max[\widetilde{x},0].$$

Gross sell orders

$$\tilde{S}=\tilde{G}-\tilde{B}\quad\text{so}\quad\tilde{B}-\tilde{S}=2\tilde{B}-\tilde{G},$$

$$\begin{aligned}\tilde{B}-\tilde{S}&=\sum_{i=1}^{N_u}2\max[\widetilde{u}_i,0]+2\max[\widetilde{x},0]-\sum_{i=1}^{N_u}|\widetilde{u}_i|+|\widetilde{x}|\\&=\sum_{i=1}^{N_u}2\max[\widetilde{u}_i,0]-|\widetilde{u}_i|+2\max[\widetilde{x},0]-|\widetilde{x}|\\&=\sum_{i=1}^{N_u}\widetilde{u}_i+\widetilde{x}=\widetilde{u}+\widetilde{x}.\end{aligned}$$

Result 2. The expected aggregate realized (half) spread without preferencing is zero.

Proof.

Aggregate realized (half) spread

$$\begin{aligned}&=\tilde{B}([p_0+\lambda(\widetilde{x}+\widetilde{u})]-\widetilde{v})+\tilde{S}(\widetilde{v}-[p_0+\lambda(\widetilde{x}+\widetilde{u})])\\&=(\tilde{B}-\tilde{S})([p_0+\lambda(\widetilde{x}+\widetilde{u})]-\widetilde{v})\\&=(\widetilde{x}+\widetilde{u})(p_0+\lambda(\widetilde{x}+\widetilde{u})-\widetilde{v}),\end{aligned}$$

$$\begin{aligned}\mathrm{E}\left[(\tilde{x}+\tilde{u})(p_0+\lambda(\tilde{x}+\tilde{u})-\tilde{v})\right]&=\mathrm{E}\left\{\mathrm{E}\left[(\tilde{x}+\tilde{u})(p_0+\lambda(\tilde{x}+\tilde{u})-\tilde{v})|\tilde{x}+\tilde{u}\right]\right\}\\&=\mathrm{E}\left\{(\tilde{x}+\tilde{u})\mathrm{E}\left[(p_0+\lambda(\tilde{x}+\tilde{u})-\tilde{v})|\tilde{x}+\tilde{u}\right]\right\}\\&=\mathrm{E}[(\tilde{x}+\tilde{u})\cdot 0]=0.\end{aligned}$$

Result 3.

$$\frac{1}{2}\frac{\mathrm{d}mmp}{\mathrm{d}f}+\frac{\mathrm{d}itp}{\mathrm{d}f}>0.$$

Proof.

$$\begin{aligned}mmp&=f\sigma_u\sqrt{\Sigma_0}\Big/\left(1\Big/\sqrt{N_u}+\sqrt{(1-f)}\right)\\&\Rightarrow\frac{\mathrm{d}mmp}{\mathrm{d}f}=\sigma_u\sqrt{\Sigma_0}\left[\frac{1/\sqrt{N_u}+\sqrt{(1-f)}+\frac{1}{2}f/\sqrt{(1-f)}}{\left(1/\sqrt{N_u}+\sqrt{(1-f)}\right)^2}\right].\end{aligned}$$

$$itp = \frac{1}{2}\sqrt{(1-f)}\sigma_u\sqrt{\Sigma_0} \Rightarrow \frac{\mathrm{d}itp}{\mathrm{d}f} = -\frac{1}{4}\sigma_u\sqrt{\Sigma_0}\frac{1}{\sqrt{(1-f)}}$$

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}mmp}{\mathrm{d}f} + \frac{\mathrm{d}itp}{\mathrm{d}f} \\
&\quad = \frac{1}{4}\sigma_u\sqrt{\Sigma_0}\left[\frac{2/\sqrt{N_u} + 2\sqrt{(1-f)} + f/\sqrt{(1-f)} - 1/(N_u\sqrt{(1-f)}) - 2/\sqrt{N_u} - \sqrt{(1-f)}}{\left(1/\sqrt{N_u} + \sqrt{(1-f)}\right)^2}\right] \\
&\quad = \frac{1}{4}\sigma_u\sqrt{\Sigma_0}\left[\frac{\sqrt{(1-f)} + f/\sqrt{(1-f)} - 1/\left(N_u\sqrt{(1-f)}\right)}{\left(1/\sqrt{N_u} + \sqrt{(1-f)}\right)^2}\right] \\
&\quad = \frac{1}{4}\sigma_u\sqrt{\Sigma_0}\left[\frac{N_u(1-f) + N_u f - 1}{N_u\sqrt{(1-f)}\left(1/\sqrt{N_u} + \sqrt{(1-f)}\right)^2}\right] \\
&\quad = \frac{1}{4}\sigma_u\sqrt{\Sigma_0}\left[\frac{N_u - 1}{N_u\sqrt{(1-f)}\left(1/\sqrt{N_u} + \sqrt{(1-f)}\right)^2}\right] > 0.
\end{aligned}$$

Appendix B. Bias adjustment

The long-run share-weighted average effective and realized spreads, $\overline{es}^p$ and $\overline{rs}^p$, are consistently estimated using the ratios of the aggregate effective and realized spread to aggregate volume. Similarly, the expected value of the preferencing measure is consistently estimated using consistent estimates for $\overline{es}^p$ and $\overline{rs}^p$. In finite samples, however, estimation error for the quantity in the denominator of the ratio leads to an upward bias. This bias can be attenuated (exacerbated) by positive (negative) correlation between the estimation errors in the numerator and denominator.

Procedures for dealing with this potential small-sample bias must make distributional assumptions for the estimation error. To see why, suppose the distribution of the sampling error results in a positive probability that the estimate in the denominator is equal to (or nearly equal to) zero, causing the ratio of the estimates to explode. Below, we make a simple assumption for the distribution of the sampling error in the denominator, and then derive a closed-form expression for the bias that results from using the ratio of estimates.

We are interested in estimating the ratio $\frac{x}{y}$ given estimates of x and y, denoted $\hat{x}$ and $\hat{y}$:

Assumption 1. $\hat{x}$ and $\hat{y}$ are unbiased, that is

$$\hat{x} = x + \varepsilon_x,\ \hat{y} = y + \varepsilon_y,\ \mathrm{E}[\varepsilon_x] = \mathrm{E}\left[\varepsilon_y\right] = 0$$

Assumption 2. The estimation error in $\hat{y}$ has two equally probable outcomes

$$\varepsilon_y = \begin{array}{ll} +\sigma_y & \text{with probability .5,} \\ -\sigma_y & \text{with probability .5.} \end{array}$$

Result.

$$\mathrm{E}\left[\frac{\hat{x}}{\hat{y}}\right] = \frac{x}{y} + \frac{\sigma_y^2(x/y) - \rho_{x,y}\sigma_x\sigma_y}{y^2 - \sigma_y^2},$$

where

$$\sigma_x^2 = \mathrm{E}\left[\varepsilon_x^2\right], \quad \sigma_y^2 = \mathrm{E}\left[\varepsilon_y^2\right], \quad \rho_{x,y} = \mathrm{E}\left[\varepsilon_x\varepsilon_y\right]/\sigma_x\sigma_y.$$

Proof. First, note that:

$$\mathrm{E}[\varepsilon_x] = (0.5)\mathrm{E}\left[\varepsilon_x|\varepsilon_y = +\sigma_y\right] + (0.5)\mathrm{E}\left[\varepsilon_x|\varepsilon_y = -\sigma_y\right] = 0$$

so

$$\mathrm{E}\left[\varepsilon_x|\varepsilon_y = +\sigma_y\right] = -\mathrm{E}\left[\varepsilon_x|\varepsilon_y = -\sigma_y\right].$$

Using the fact that $\mathrm{E}[\varepsilon_x] = \mathrm{E}\left[\varepsilon_y\right] = 0$,

$$\begin{aligned} \rho_{x,y}\sigma_x\sigma_y &= \mathrm{E}\left[\varepsilon_x\varepsilon_y\right] \\ &= (0.5)\mathrm{E}\left[\varepsilon_x\varepsilon_y|\varepsilon_y = +\sigma_y\right] + (0.5)\mathrm{E}\left[\varepsilon_x\varepsilon_y|\varepsilon_y = -\sigma_y\right] \\ &= (0.5)\sigma_y\mathrm{E}\left[\varepsilon_x|\varepsilon_y = +\sigma_y\right] - (0.5)\sigma_y\mathrm{E}\left[\varepsilon_x|\varepsilon_y = -\sigma_y\right] \\ &= \sigma_y\mathrm{E}\left[\varepsilon_x|\varepsilon_y = +\sigma_y\right] \end{aligned}$$

so

$$\mathrm{E}\left[\varepsilon_x|\varepsilon_y = +\sigma_y\right] = \rho_{x,y}\sigma_x \quad \text{and} \quad \mathrm{E}\left[\varepsilon_x|\varepsilon_y = -\sigma_y\right] = -\rho_{x,y}\sigma_x.$$

Taking the conditional expectation of $\hat{x}/\hat{y}$ gives

$$\begin{aligned} \mathrm{E}\left[\frac{\hat{x}}{\hat{y}}\right] &= (0.5)\mathrm{E}\left[\frac{\hat{x}}{\hat{y}}|\varepsilon_y = +\sigma_y\right] + (0.5)\mathrm{E}\left[\frac{\hat{x}}{\hat{y}}|\varepsilon_y = -\sigma_y\right] \\ &= (0.5)\left[\frac{x + \rho_{x,y}\sigma_x}{y + \sigma_y}\right] + (0.5)\left[\frac{x - \rho_{x,y}\sigma_x}{y - \sigma_y}\right] \\ &= \frac{xy - \rho_{x,y}\sigma_x\sigma_y}{\left(y + \sigma_y\right)\left(y - \sigma_y\right)} \\ &= \frac{x}{y} + \frac{\sigma_y^2(x/y) - \rho_{x,y}\sigma_x\sigma_y}{y^2 - \sigma_y^2}. \end{aligned}$$

Thus, our bias-adjusted estimate for

$$\frac{x}{y} \quad \text{is} \quad \frac{\hat{x}}{\hat{y}} - \frac{\hat{\sigma}_y^2(\hat{x}/\hat{y}) - \hat{\rho}_{x,y}\hat{\sigma}_x\hat{\sigma}_y}{\hat{y}^2 - \hat{\sigma}_y^2}.$$

We use the six-month time series for each firm to estimate $\hat{\sigma}_x$ (gross effective and realized spread), $\hat{\sigma}_y$ (gross volume), and $\hat{\rho}_{x,y}$, and then use these to produce bias-adjusted estimates for $\overline{es}^p$ and $\overline{rs}^p$. We then use the ratio of the bias-adjusted estimates for $\overline{es}^p$ and $\overline{rs}^p$, along with an adjustment based on estimates of the standard deviations and correlation in sampling errors based on the six-month time series for each firm, to produce a biased-adjusted estimate for our preferencing measure *pm*. The results (not shown) are very similar to those shown in Table 3.

References

Affleck-Graves, J., Hegde, S., Miller, R., 1994. Trading mechanisms and the components of the bid-ask spread. Journal of Finance 49, 1471–1488.

Battalio, R., 1997. Third market broker-dealers: cost competitors or cream skimmers? Journal of Finance 52, 341–352.

Battalio, R., Holden, C., 2001. A simple model of payment for order flow, internalization, and total trading cost. Journal of Financial Markets 4, 33–71.

Benveniste, L., Marcus, A., Wilhelm, W., 1992. What's special about the specialist? Journal of Financial Economics 32, 61–86.

Bessembinder, H., 1999. Trade execution costs on NASDAQ and the NYSE: a post-reform comparison. Journal of Financial and Quantitative Analysis 34, 387–408.

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

Bessembinder, H., 2003b. Quote-based competition and trade execution costs in NYSE-listed securities. Journal of Financial Economics 70, 385–422.

Bessembinder, H., Kaufman, H., 1997. A comparison of trade execution costs for NYSE and NASDAQ-listed stocks. Journal of Financial and Quantitative Analysis 32, 287–310.

Boehmer, E., 2005. Dimensions of execution quality: recent evidence for U.S. equity markets. Journal of Financial Economics 78, 553–582.

Chakravarty, S., Sarkar, A., 2002. A model of broker's trading, with applications to order flow internalization. Review of Financial Economics 11, 19–36.

Chordia, T., Subrahmanyam, A., 1995. Market making, the tick size, and payment-for-order flow: theory and evidence. Journal of Business 68, 543–575.

Chung, K., Chuwonganant, C., McCormick, D.T., 2004. Order preferencing and market quality on NASDAQ before and after decimalization. Journal of Financial Economics 71, 581–612.

Easley, D., Kiefer, N., O'Hara, M., 1996. Cream-skimming or profit sharing? The curious role of purchased order flow. Journal of Finance 51, 811–834.

Hansch, O., Naik, N., Viswanathan, S., 1999. Preferencing, internalization, best execution, and dealer profits. Journal of Finance 54, 1799–1828.

Hasbrouck, J., Saar, G., 2002. Limit orders and volatility in a hybrid market: the Island ECN. New York University working paper.

Huang, R., Stoll, H., 1996. Dealer versus auction markets: a paired comparison of execution costs on NASDAQ and the NYSE. Journal of Financial Economics 41, 313–357.

Jones, C., Lipson, M., 2003. Why NYSE retail orders get better (yes, better) executions, Columbia University and University of Virginia working paper.

Kandel, E., Marx, L., 1999. Payments for order flow on NASDAQ. Journal of Finance 54, 35–66.

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

Lin, J-C., Sanger, G., Booth, G., 1998. External information costs and the adverse selection problem: a comparison of NASDAQ and NYSE stocks. International Review of Financial Analysis 7, 113–136.

Lipson, M., 2004. Competition among market centers. University of Virginia working paper.

Parlour, C., Rajan, U., 2003. Payment for order flow. Journal of Financial Economics 68, 379–411.

Parlour, C., Seppi, D., 2003. Liquidity-based competition for order flow. Review of Financial Studies 16, 301–343.

Peterson, M., Sirri, E., 2003. Order preferencing and market quality on U.S. equity exchanges. Review of Financial Studies 16, 385–415.

U.S. Securities and Exchange Commission, 1997. Report on the Practice of Preferencing, Washington, DC.

U.S. Securities and Exchange Commission, 2001. Report on the Comparison of Order Execution Across Equity Market Structures, Washington, DC.

Weston, J., 2000. Competition on the NASDAQ and the impact of recent market reforms. Journal of Finance 55, 2565–2598.