June 19, 2023

TO:

Vanessa Countryman, Secretary
U.S. Securities and Exchange Commission
100 F Street, NE
Washington, DC 20549-1090

FROM:

S.P. Kothari
Gordon Y Billard Professor of Accounting and Finance
MIT Sloan School of Management
100 Main Street
Cambridge, MA 02142

FROM:

Travis L. Johnson
Associate Professor of Finance
The University of Texas at Austin
2110 Speedway Stop B6600
Austin, TX 78751

RE: Request for Comment on the Proposed Best Execution Rule (File Number S7-32-22)

Dear Ms. Countryman,

We appreciate the opportunity to comment on the proposed Best Execution rule ("the Proposal"). We also submitted a substantively-identical comment on the proposed Order Competition rule (File Number S7-31-22), which relies on some of the same economic analysis.

Our comment represents our opinion and not necessarily that of Robinhood Markets Inc., with whom we have a financial relationship, or Massachusetts Institute of Technology and The University of Texas at Austin, our respective employers.

1. Executive Summary

We limit our attention in this comment to the Proposal's economic analysis on the relation between payment for order flow ("PFOF") and the prices received by retail trades ("execution quality"). The Proposal's analysis is fundamentally flawed. It makes dubious assumptions, is missing important control variables, overstates the number of independent observations, and has miniscule economic significance.

2. **Summary of Proposal's Analysis**

We begin with an accessible explanation of how the Proposal uses a statistical analysis to conclude that "wholesalers provide worse execution quality to brokers that receive more PFOF."[1] This conclusion is relevant for the best execution rule, the Proposal argues, because it suggests that retail traders could be receiving better execution prices.

The ideal experiment for determining how PFOF affects execution quality would be to compare trades with smaller PFOF to otherwise-identical trades where the broker received larger PFOF. Assuring that the samples in the analysis are otherwise identical is important because otherwise any differences in execution quality could be driven by the differences unrelated to PFOF.[2] For example, if the orders with larger PFOF tended to be for smaller-cap stocks with less volume, these orders will have larger trading costs due to the reduced liquidity rather than the PFOF itself. In the absence of "otherwise identical" samples, a statistical regression with the proper data and control variables can potentially be interpreted as if the samples are identical.

A. Simple Comparison of PFOF and Non-PFOF Brokers (Table 15)

With this ideal experiment in mind, we evaluate the Proposal's analysis. The Proposal's analysis begins by dividing all retail brokers into two buckets, those that received PFOF from wholesalers and those that did not. Table 15 shows that retail trades at non-PFOF brokers receive better prices (i.e., lower prices for buy orders and higher prices for sells) than trades at PFOF brokers. It does so using four measures for execution quality gleaned from the consolidated audit trail ("CAT"):

1) *Effective Spread*, defined as the execution price minus the midpoint of the national best bid ("bid") and national best offer ("ask") prices at the time of the order for buy orders, and the midpoint minus the execution price for sell orders, both scaled by the average share price. This captures how much the trader loses relative to trading at the midpoint, a benchmark for the fair market price. Higher values indicate worse execution quality.[3]
2) *E/Q,* the ratio of *Effective Spread* to the average *Quoted Spread*, defined as half the difference between ask and bid prices, scaled by the average share price. This captures

[1] See page 242 of the Proposal.
[2] See Section 3.2 of Angrist and Pischke (2009) for a discussion of causal inference with potential omitted variables.
[3] See Goynko, Holden, and Trzcinka (2009) for a comparison of this with other trading cost measures.

what fraction of the bid-ask spread the retail customer actually pays. Higher values indicate worse execution quality.

3) *Realized Spread*, defined as the execution price minus the midpoint five minutes after the trade for buy orders, and the midpoint five minutes after the trade minus the execution price for sell orders, both scaled by the average share price. This captures how much the trader loses relative to trading at the future midpoint. Higher values indicate worse execution quality.
4) *Price Improvement*, defined as the ask minus the execution price for buy orders, and the execution price minus the bid for sell orders, both scaled by the average share price. This captures how much the trader gains relative to trading at the bid or ask prices. Higher values indicate better execution quality.

Table 15 shows that *Effective Spreads* and *E/Q* are slightly higher, while *Realized Spreads* and *Price Improvement* are slightly lower, for trades made via PFOF brokers than trades made via non-PFOF brokers. Three of these results indicate worse execution quality at PFOF brokers while one (*Realized Spreads*) suggests better execution quality.

While this exercise is easy to understand, the differences cannot unambiguously be interpreted as being attributable to PFOF itself because large differences exist in the types of stocks traded by customers of Non-PFOF and PFOF brokers. Table 15 shows that share prices are lower for trades made via PFOF brokers than Non-PFOF brokers. Given brokers specialize in serving different clientele,[4] it is also likely to that other large differences in the types and quantities of shares traded between PFOF and non-PFOF brokers.

B. Regression Analysis Explaining Execution Quality using PFOF Rate (Table 16)

To control for differences in the stocks traded by customers of PFOF and Non-PFOF brokers, and to assess whether magnitude of PFOF (in addition to its existence) affects execution quality, the Proposal turns to a regression analysis. The dependent or explanatory variable is one of the above execution quality metrics from the CAT data. The regressions, presented in Table 16, have over 12-14 million observations because the CAT data provides execution quality metrics aggregated

[4] See Fong, Gallagher, and Lee (2014) for evidence on differences in trading behavior across broker types.

into buckets having the same stock, week, order type, order size category, wholesaler, and retail broker MPID. In tabular form, the data look like[5]:

Stock	Week	Order Type	Order Size	Wholesaler	Broker	E/Q	…
FB	3/7-3/11	Market	< 100	Virtu	Fidelity	0.57	…
MSFT	1/3-1/7	Market	100-499	Citadel	Schwab	0.74	…
… 12-14 million observations							

The main independent variable of interest is PFOF Rate, defined as the average PFOF amount divided by the average share price. Unfortunately, the PFOF amount is not available in the CAT data. Instead, the Proposal relies on brokers' disclosures mandated by Rule 606. These disclosures are aggregated into much broader buckets than the CAT data, lumping together all executions in each month with the same order type, wholesaler, retail broker, and S&P 500 indicator. In tabular form, the data look like[6]:

Stock Category	Month	Order Type	Wholesaler	Broker	PFOF (¢/100 shares)
S&P 500	Jan. 2022	Market	Virtu	Robinhood	38.80
Non-S&P 500	Jan. 2022	Market	Virtu	Robinhood	52.83
… Approximately 5,160 observations					

Critically, the data contain no information about how PFOF varies across specific stocks, weeks within the month, or order size buckets. Aggregating in this manner dramatically reduces the sample size of PFOF data to at most 5,160 observations[7].

To populate the 12-14 million observations in finer CAT data, the Proposal computes *PFOF Rate* for CAT observation i by dividing the *PFOF Amount* (measured in ¢/100 shares) for the corresponding stock category/month/order type/wholesaler/broker c by the volume-weighted average price ("VWAP") for the stock in that week:[8]

[5] This table illustrates the format of available data but does not contain actual data.
[6] This table illustrates the format of available data but does not contain actual data.
[7] The Rule 606 data used in the Proposal's analysis spans 3 months, 43 broker-dealers (See Table 2 of the Proposal), 4 order-type categories, and 2 stock categories. Assuming each broker-dealer used 5 different wholesalers for each category, more than the typical values we find in Rule 606 reports, the total sample size for PFOF is at most $3 \times 43 \times 4 \times 2 \times 5 = 5{,}160$ observations.
[8] See Footnote 460 of the Proposal.

$$PFOF\ Rate_i = \frac{PFOF\ Amount_c}{VWAP_i}$$

In doing so, the Proposal replicates the same *PFOF Amount* across the thousands of stocks and order sizes within the two broad categories (S&P 500 and non-S&P 500). The only variation across these thousands of observations comes from the share price $VWAP_i$.

The regression results in Table 16 indicate that higher *PFOF Rate* predicts higher *E/Q*, *Effective Spread*, and *Realized Spread,* and lower *Price Improvement*. All of these suggest worse execution quality (higher costs) when *PFOF Rate* is larger.

3. Flaws in the Proposal's Methodology

We focus our discussion of the flaws in the Proposal's methodology on the regression analysis in Table 16. The regression approach is more promising than the simple differences in means presented in Table 15 because it attempts to control for the large differences in the types of stocks traded by customers of Non-PFOF and PFOF brokers.

Unfortunately, four major flaws remain in the regression analysis. Three of these arise from a core limitation of the Rule 606 data: **the Proposal's data has no variation across stocks in PFOF amounts**, and instead the analysis must rely only on the variation across broker/wholesaler pairs. This limitation is fatal in the aforementioned ideal experiment because there is no way of knowing or controlling for the possibility that variation in PFOF amounts is attributable to variation in the stocks and quantities traded rather than the PFOF amounts for the same stock/quantities.

A. Dubious Assumptions

Absent data on variation across stocks and quantities in PFOF amounts, the Proposal assumes that *PFOF Amount* is **the same across all stock/quantities** within each grouping of the Rule 606 data. This is highly unlikely, as illustrated by the stark differences in average PFOF payments between S&P 500 and non-S&P 500 stocks, and across order types, shown in Table 2 of the Proposal. If such large differences exist between these groups, differences within each group, for example as a function of order size or share price, are likely to be as large or larger.[9]

If PFOF amounts exhibit unobserved variation across stocks/quantities, the Proposal's analysis could still be unbiased if the variation is uncorrelated with execution quality measures.[10] This too is highly unlikely, as variation in PFOF amounts is expected to be driven by stock characteristics such as liquidity that are also correlated with the execution quality measures.[11] Instead, estimates in Table 16 of the proposal are likely to be biased because some omitted stock characteristic explains both the average PFOF received by broker/wholesaler pairs and execution quality metrics.

A related assumption is that when PFOF amounts data are missing, the Proposal assumes the broker receives 20¢/100 shares.[12] There is no reason to believe this is accurate. A standard approach in academic research would be to drop observations for which PFOF amounts were missing.[13]

B. Missing Important Control Variables

The Proposal uses control variables to address the possibility that variation in execution quality metrics is driven by stock characteristics reflected in *PFOF Rate*, instead of *PFOF Rate* itself affecting execution quality. However, important control variables are missing from the analysis. Given that *PFOF Rate* only varies across stocks within each Rule 606 bucket due to variations in the $VWAP$ scaling variable, an important control variable is $\frac{1}{VWAP_i}$. This variable, unlike the $VWAP_i$ variable the Proposal does include, would control for any variation driven purely by the denominator of *PFOF Rate.*

The omission of $\frac{1}{VWAP_i}$ as a control variable is particularly likely to introduce a bias because the dependent variables are also likely related to share prices. Lower share price stocks (higher $\frac{1}{VWAP_i}$)

[9] No public data exists on how PFOF amounts vary across stocks or trade sizes. However, all available evidence on trading costs, e.g., in Frazzini, Israel, and Moskowitz (2012), indicates large differences across stocks and as a function of trade size.
[10] See Section 3.2 of Angrist and Pischke (2009) for technical details.
[11] See McInish and Wood (1992) and related work for evidence on how bid-ask spreads vary as a function of firm characteristics.
[12] See Footnote 460 of the Proposal.
[13] See Ernst and Spatt (2022) for an example of research using Rule 606 reports data that excludes missing data from the sample rather than assuming specific values.

are likely to have worse execution quality, consistent with the coefficients presented in Table 16. This prediction holds despite the latter three dependent variables also being scaled by *VWAP* because spreads are larger relative to prices for lower share price stocks.[14]

Average order size, short interest, and return volatility are also relevant control variables the Proposal omits. Each has a well-established effect on market makers and so is likely related to both execution quality measures and PFOF Rate.[15]

C. Overstates the Number of Independent Observations

Regression analysis assesses how likely it is that an apparent relation in the data is due to random chance. If it is very unlikely, we say that the relation is "statistically significant," and otherwise we say it is "statistically insignificant." The calculation of statistical significance depends critically on sample size because random variation matters less and less as the sample size increases.

The relations between *PFOF Rate* and execution quality measures in Table 16 are all statistically significant in part because the regressions have 12-14 million observations. However, these observations rely on at most 5,160 distinct observations of the PFOF amounts, meaning **the sample size is effectively 2600 times smaller than the Proposal assumes**.

There are two approaches to correcting statistical significance measures for this dearth of independent variation.[16] The first would be to use "cluster-robust standard errors" to allow for observations within the same Rule 606 bucket to be correlated instead of independent.[17] The second would be to collapse the CAT data to averages within each Rule 606 bucket and run the regression on the much smaller, but more accurate to the actual variation in the data, sample of at most 5,160 observations with different PFOF amounts.[18]

---

[14] See Conroy, Harris, and Benet (1990).
[15] See Frazzini, Israel, and Moskowitz (2012), Zhoa, Cheng, and Cheng (2013), and Li and Wu (2006) for evidence on the relations between bid-ask spreads and trade size, short interest, and return volatility, respectively.
[16] See Section 8.2 of Angrist and Pischke (2009).
[17] The Proposal instead clusters standard errors at the stock level, an important but insufficient correction. Clustering by both stock and Rule 606 bucket would be more accurate.
[18] The proposal includes wholesaler, order size category, and stock fixed effects as additional control variables. These absorb any variation in the independent and dependent variables attributable purely to wholesalers, order size categories, or stocks. This further reduces the amount of independent variation and does not address the standard error issue we analyze here.

The Proposal does not employ either of these approaches. If it did, the standard errors would substantially increase, possibly rendering the results statistically insignificant.[19]

D. Small Economic Significance

Even if the identified biases in the regression results were to be ignored, a simple calculation reveals that the purported impact of PFOF on retail traders' costs is economically small. The average order gets \$0.0013 per share in PFOF (Table 2). The average share price for orders executed by wholesalers in the sample is \$29.87 (Table 6). This means that the average *PFOF Rate* is $\frac{0.0013}{29.87} = 0.435bp$. The estimates in Table 16 therefore suggest that if all of PFOF were to be eliminated, it would result in the following effects:

Measure	Predicted Effect of Removing all PFOF	PFOF Sample Average (Table 15)
E/Q	$-0.0132 \times 0.435 = -0.0057$	0.37
Effective Spread (bp)	$-0.217 \times 0.435 = -0.0944$	1.86
Realized Spread (bp)	$-0.211 \times 0.435 = -0.0918$	0.85
Price Improvement (bp)	$0.170 \times 0.435 = 0.0740$	2.34

The predicted effects in the second column are miniscule compared to the sample averages of the respective dependent variables in the third column. This indicates PFOF rates are at best a second-order determinant of execution quality for retail traders.

4. References

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge University Press.

Conroy, R. M., Harris, R. S., & Benet, B. A. (1990). The effects of stock splits on bid-ask spreads. *The Journal of Finance*, *45*(4), 1285-1295.

[19] Standard errors are proportional to $1/\sqrt{T}$, where T is the sample size (see, e.g. Section 4.4.4 of Cameron and Trivedi (2005). All else equal, a reduction in sample size by a factor of 2600 could therefore reduce standard errors by a factor of 51, rendering the effects in Table 15 insignificant.

Ernst, T., & Spatt, C. S. (2022). *Payment for order flow and asset choice* (No. w29883). National Bureau of Economic Research.

Fong, K. Y., Gallagher, D. R., & Lee, A. D. (2014). Individual investors and broker types. *Journal of Financial and Quantitative Analysis*, *49*(2), 431-451.

Frazzini, A., Israel, R., & Moskowitz, T. J. (2012). Trading costs of asset pricing anomalies. *Fama-Miller Working Paper, Chicago Booth Research Paper*, (14-05).

Goyenko, R. Y., Holden, C. W., & Trzcinka, C. A. (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics*, *92*(2), 153-181.

Li, J., & Wu, C. (2006). Daily return volatility, bid-ask spreads, and information flow: Analyzing the information content of volume. *The Journal of Business*, *79*(5), 2697-2739.

McInish, T. H., & Wood, R. A. (1992). An analysis of intraday patterns in bid/ask spreads for NYSE stocks. *The Journal of Finance*, *47*(2), 753-764.

Zhao, Y., Cheng, L. Y., Chang, C. C., & Ni, C. Y. (2013). Short sales, margin purchases and bid–ask spreads. *Pacific-Basin Finance Journal*, *24*, 199-220.