MONASH University

Does Imposing Trades With Price Improvement Take Liquidity Away from Dark Pool?

Bagas Trilaksonoaji
32687044

Associate Professor Paul Lajbcygier
Project Supervisor

Report for
ETC5543 - Business Analytics Creative Activity

28 October 2024

MONASH
BUSINESS
SCHOOL

Department of Econometrics & Business Statistics

baga0002@student.monash.edu

Table of contents

1 Abstract

The effect of the Trade With Price Improvement rule, which went into effect on 26 May 2013, on `at-the-spread` crossings in the `dark pool` of the Australian Securities Exchange (ASX), is examined in this study. The drop in the proportion of trading volume from `at-the-spread` crossings from 52% to 1% following the rule change indicates a notable shift in trading volume. Although the breaching of the rule initially accounted for only about 5% of all internal crossings, monthly statistics showed a steady increase in rule violations, peaking at 12% in December 2015. The majority of rule violations were done by only two brokers. We also use machine learning models to understand violation trends. The result emphasized the importance of market conditions, namely market spread and market depth, as the main contributors to the violation of the TWPI rule.

Link to GitRepo: https://github.com/bgstrlk/ETC5543

2 Introduction

There were numerous equities exchanges established in the long history of Australia. Almost every big city in Australia once had their own stock exchange, namely Sydney, Brisbane, Melbourne, Adelaide, Perth, and even Hobart. The Australian Parliament merged all six stock exchanges in 1987, creating the Australian Stock Exchange (Australian Stock Exchange, n.d.). Two decades later, this equities exchange merged with Sydney Futures Exchange to become the Australian Securities Exchange. Although its official name is now ASX Limited, it is more convenient to simply refer to it as ASX. In this report, however, we will use the term ASX to refer to the company or the equities market.

As a market operator, ASX's main tasks include reviewing listing applications, supervising listed entities as well as listed brokers, operating and maintaining an electronic trading platform for buying and selling stocks, and providing clearing and settlement services. (Australian Stock Exchange, n.d.).

ASX operates under the regulatory framework of two government agencies, the Australia Securities and Investments Commission (ASIC) and the Reserve Bank of Australia (RBA) (Australian Stock Exchange, n.d.). ASIC oversees stock market trading and enforces financial market conduct laws, while the Reserve Bank of Australia (RBA) oversees clearing and settlement-related services.

2.1 Background and Motivation

Anyone who wants to buy or sell stocks can place an order on the ASX trading platform. When the price of the buy and sell side matches, the trade can be executed. Outstanding orders will remain on the platform and visible to market participants. This mechanism promotes transparency and market equilibrium in supply and demand, thereby ensuring fair stock price formation (Australian Securities and Investments Commission, 2022). That is why we often call this trading venue provided by the ASX trading platform `lit market` or `lit pool`.

On the other hand, there are other venues that do not disclose the orders placed, at least until the transaction is executed. We will use the term `dark pool` to refer to this kind of trading venue. There are several `dark pools` available for stock trading in Australia. However, in this project, we will focus on the `dark pool` executed on the broker's internal crossing system. This kind of `dark pool` is often called NBBO trade, which stands for National Best Bid and Offer Prices (Charles Lane Advisory Pty Ltd, 2014).

Crossing system is an electronic platform provided by the brokers to match orders placed by their clients with their other clients or with their own account (Australian Securities and Investments Commission, 2014). In other words, orders can be matched on the broker's internal system before they are passed through to the ASX trading platform. For the next section, we will use the term `crossing` to refer to the trades in the broker's internal crossing system.

What is the benefit for brokers and traders if they trade on the crossing system instead of the ASX `lit pool`? One of the primary benefits is the significantly lower trading fees. Trades on the ASX platform are charged about 0.15 basis points per trade, whereas internal crossing charges as much as 0.04 basis points per trade, capped at $1,000 per month (Australian Securities Exchange, 2024)). The difference is quite huge given how much a broker can trade on a daily basis.

It is important to note that the rule in which internal crossings referred to was differed before and after 26 May 2013:

1. **Before**

   Rule 4.2.3(1)(b) of the ASIC Market Integrity Rules (Competition in Exchange Markets) 2011, regulates the `At or Within the Spread` trades. As the name suggests, the executed trade price must be at the same as the best available bid or best available offer of the `lit market` (at-the-spread), or better (within-the-spread). This implies that the price executed in this "dark pool" cannot be lower than the best bid price or higher than the best offer price in the "lit market".

2. **After**

   The rule had changed, based on Rule 4.2.3 of the ASIC Market Integrity Rules (Competition in Exchange Markets) Amendment 2012 (No. 1), the trades on the crossing system must provide meaningful price improvement (Australian Securities and Investments Commission, 2014).

> Trades at the best bid or best offer price of the `lit market` are not in the option anymore. This rule was effective on 26 May 2013. ASIC and ASX refer to this rule as `Trades With Price Improvement`, or next we can refer to this as TWPI. This study will specifically focus on the internal crossings, which follow the TWPI rule.

The amendment of the rule in May 2013 was to address the negative impact of the dark trades on the price formation, specifically from the `at-the-spread` trades (Australian Securities and Investments Commission, 2014), due to traders may prefer to do the trades on the `dark pool` as it costs less fees compared to the `lit pool`. The adverse effect on price discovery was one of the main reasons why the regulator wanted more trading to occur in the `lit pool`.

On the other hand, the ASX was also motivated by an economic incentive. On October 14, 1998, ASX became a publicly owned company after demutualising itself. The main analysis behind this was that "*ASX needed to become more flexible, responsive, and commercially focused, capable of quickly taking up emerging commercial opportunities*" (Australian Securities and Investments Commission, 2004). Therefore, ASX, being an all-for-profit enterprise, would seize any business opportunity to increase its revenue and maximize its returns to shareholders. ASIC amended the rules to provide meaningful price improvements, which would shift a significant portion of dark trades to the ASX platform, thereby increasing its profit from trading fees.

Key Motivation

In May 2022, ASIC published a Market Integrity Update (Issue 137), which included a review of TWPI. In the article, ASIC mentioned that several participants were reporting a significant number of their TWPIs without any price improvement. ASIC emphasized that such trades must be done in the `lit market` because they will promote liquidity and price formation. TWPI is an exception because it offers better prices for the investor—a price that the `lit market` cannot provide.

Following up on the Market Integrity Update, ASIC issued an infringement notice to Wilson Advisory and Stockbroking Pty Ltd, requiring them to pay $548,328 to ASIC on behalf of the Commonwealth. The notice was given on 16 December 2022, while published on an ASIC media release on 3 February 2023.

The penalty shows that ASIC has started to take the breach of the TWPI rule more seriously. Probably due to the increasing number of dark trades that should be done on the `lit market`, as ASIC intention of rule change in 2013 was to increasing the proportion of trading on the `lit market`, hence, enhancing the fairness on the market price formation Australian Securities and Investments Commission, 2014

Based on these insights, I'm intrigued to analyse the `at-the-spread` dark trades before and after the rule amendment in 26 May 2013. Different from the ASIC report (Report 394) on May 2014, this study will focus on the internal crossings mechanism, among all other dark pool mechanisms. Furthermore, the internal crossings violation on the TWPI rule will be analysed, along with the model that explains the pattern of the violation.

2.2 Purpose and Key Questions

The main objective of this study is to look into how the rule change in 2013 affected `at-the-spread` crossings reported as TWPI. The way brokers act and trade on the `dark pool` should be greatly affected by this change, as they must move the `at-the-spread` crossings to the `lit pool`. Moreover, this study will also examine the violation of the TWPI after the rule change.

The study will compare transaction data before and after the 2013 regulation change, with a focus on `at-the-spread` crossings. Trade volumes, prices, time stamps, broker IDs, and several other variables will be included in the analysis. There will be simple statistical methods used to compare trade distributions, and machine learning methods may be used to find trends of the violations in the TWPI rule.

Understand the Impact on At-the-Spread Crossings

`At-the-spread` crossings were possible before the rule change, which meant that trades could be made at the best bid or best offer price available on the `lit market`. However, the 2013 amendment says that trades must lead to price improvements, in other words, `at-the-spread` crossings are no longer allowed. The main question to ask is the proportion of `at-the-spread` crossings among all TWPI-reported trades before and after 26 May 2013. This comparison will help to figure out whether the rule change stopped people from doing `at the spread` crossings at the `dark pool`. Furthermore, the dollar value that possibly moved to the `lit market` from the `dark pool` will also be analysed.

Key Question:

1. Did the number of `at-the-spread` crossings drop significantly after 26 May 2013?
2. How much was approximately the value of `at-the-spread` internal crossings that moved to the lit market after the rule change?

Identify the Violation Pattern

ASIC's media release and publication in the last couple of years suggested that there have been TWPI rule violations in recent years. This study will examine the proportion of the violations, identify which brokers do the most violations, and find any pattern by creating a model to help spot non-compliance in the future.

Key Question:

1. How many proportion of violations occurred following the rule change?
2. What kind of brokers violate the rules the most? Are some brokers more likely to break TWPI rules than others?
3. What conditions contributed to the occurrence of the violations?

2.3 Significance and Contributions

There are three main areas in which this study may contribute: academic, regulatory, and economic.

Economic Significance

The purpose of the rule amendment in 2013 was to encourage meaningful price improvement on the `dark pool` (ASIC, 2014). Without price improvement, trading must be done on the ASX trading platform to promote liquidity on the market. Charles Lane Advisory Pty Ltd report (Attachment to ASIC Report 394) in May 2014 has thoroughly analyse the impact of this rule change and stated that there was a decline in the dollar value from $54.9 billion to $21 billion, 100 days prior the rule change and after. Our study aims to validate this figure by focusing only on internal crossings and extending the analysis over a one-year period, with the goal of providing a more in-depth understanding of the rule's economic impacts. From a financial point of view, it is also essential to comprehend TWPI violations since they have the potential to compromise market integrity in general (Australian Securities and Investments Commission, 2014).

Regulatory Significance

Understanding the magnitude of the TWPI violation since its amendment in 2013 may provide regulators with new insights to address the TWPI reporting issues at the prevention level rather than through corrective action. Furthermore, if the detection model developed in this study proves effective, regulators might utilize it as part of an automated enforcement mechanism to ensure TWPI reporting follows the regulations.

Academic Significance

There is a lot of academic literature on special orders and hidden orders, but this study adds a new dimension by focusing on TWPI trades. This study's novel contribution is the use of a prediction model to identify patterns and conditions that result in TWPI rule violations. The study might help academics learn more about how rules are enforced in financial markets and the situations where traders are most likely to take advantage of the system loopholes.

3 Data and Methodology

This study use transaction and market data provided by The Securities Industry Research Center of Asia-Pacific (SIRCA). Founded in 1997, SIRCA is a non-profit organization mostly consisting of academics from Australian and New Zealand universities. It contains a variety of financial market data, including stock exchanges data, from Australia, New Zealand, and even the United States (SIRCA, n.d.). SIRCA members are free to gather data for research purposes, provided that the university they are at have paid subscription fee to SIRCA.

The ASX transaction data provided by SIRCA consists of the stock code, date and time, price, volume, dollar value, broker ID, and other related variables. In addition, the ASX market data includes the best bid and best offer order prices, as well as the volume of those orders. These best bid and best offer prices are used to calculate the spread, price and volume, at each particular time.

We use datasets from stocks listed on the S&P/ASX200 index, as this index represents a group of stocks with the biggest market capitalization. This indicates that these stocks are the biggest and most valuable ones on the ASX. Market capitalization and liquidity often have a positive correlation. These traits from the S&P/ASX200 index ensure sufficient trading interest from major brokers (Duong et al., 2018), which is important for gauging the effect of the regulation change.

We will use data from the introduction of the `at or within the spread` rule in November 2011. However, since our investigation found that there were no internal crossings in 2011, we will use data from 2012 until the end of 2015.

The transaction dataset we obtained from SIRCA was filtered to include only internal crossing trades, marked as “OFFTR” in the `RecordType` variable and as “NX XT” in the `Qualifiers` variable. We combined this data with market data, allowing us to include every transactions with the best bid and offer price and volume of the market. Combining these many data requires enormous processing power, therefore, we use the help of Nectar Research Cloud, a cloud computing service from the Australia Research Data Commons.

Methodology

This study uses a methodical approach to look at how changes in regulations affect `at-the-spread` crossings in dark pools, particularly on the broker’s internal crossing system. This study also looks for trends in TWPI rule violations. With the help of SIRCA's ASX transaction and market data, we hope to comprehend the regulatory and economic effects of the 2013 rule change.

Rule Change Impact on At-the-Spread Crossings

Analyzing the proportion before and after the rule adjustment on 26 May 2013 is the first step in determining how the rule change would affect `at-the-spread` transactions. The frequency and characteristics of `at-the-spread` crossings in the internal crossing system before and after the rule change will be compared. In order to guarantee sufficient liquidity and representation of the whole market, we concentrate our analysis on the S&P/ASX200 stocks. This allows us to determine if the regulatory adjustment was successful in limiting `at-the-spread` trades to close to none.

To assess the economic implications, we evaluate the aggregate dollar value of `at-the-spread` trades over two periods: the 100 trading days preceding and the 100 trading days succeeding the rule amendment on 26 May 2013. This controlled period reflects similar transaction trends or patterns. The value-based analysis will measure the economic consequences of trades that may have shifted to the 'lit market' following the amendment, thus assessing the possible enhancement in market transparency and fairness.

Violation Pattern Analysis

We utilize logistic regression and random forest models to discover and analyze violations of TWPI rules, as both are frequently employed for binary classification and pattern recognition. These models will assist in identifying circumstances in which TWPI violations are more likely. Every trade will be evaluated for compliance to the TWPI regulations, utilizing parameters such as trade volume, timing, and spread to discern the pattern. Logistic regression yields probabilistic estimates of violations, but random forest facilitates the examination of complex variable interactions, providing both interpretability and robustness.

Statistical Measures and Tools

We will evaluate model performance utilizing balanced accuracy, specificity, and sensitivity, especially considering the apparent imbalance between complying and non-compliant trades. `R` machine learning and statistical packages, such as `tidymodels` and `caret`, will be utilized for processing, modeling, and validating outcomes in implementation and classification analysis.

4 Result and Discussion

4.1 Rule Change Impact on At-the-Spread Crossings

We compare the number of internal crossings before and after the rule change on 26 May 2013. The at-the-spread internal crossings can be done before the meaningful price improvement rule introduction; hence, only within-the-spread crossings can be done after 26 May 2013. Table 1 and Table 2 show the monthly number and proportion of internal crossings before and after the rule change, respectively. In addition, the total proportion for each trade price group (within/at/outside-the-spread) can be seen from Table 3.

From the start of the rule until the amendment, the monthly proportion of 'at-the-spread' internal crossings consistently ranged between 40% and 60%. The highest proportions were in October–December 2012, with 61%–62% of the total internal crossings. After subtracting January 2012, which had the lowest number of crossings due to the transition, April 2013 had the lowest proportion, accounting for only 39% of the total crossings.

Table 1 shows that there were a number of violations of the early rule of internal crossings. However, the proportion was not significant, consistently falling below 2%, with the exception of September 2012. The total proportion of at-the-spread crossings from the introduction of the at or within-the-spread rule until the amendment in May 2013 was 52%.

These numbers and percentages show that there were a balanced proportion of within-the-spread and at-the-spread crossings, with only an insignificant number of violations (outside-the-spread), from the introduction of the rule in November 2011 until the amendment in May 2013.

Looking at the proportion of at-the-spread crossings after the rule change, we can see from Table 2 that the percentages are constantly low under 2%, except for the first two months, June and July. (May 2013 here only accounts for 5 days). If we look at the proportion of total at-the-spread crossings from 26 May 2013 to December 2015, the percentage was only 1.23% (see Table 3).

It is also worth mentioning that the monthly frequency of within-the-spread crossings in this period was smaller than the previous period. The frequency was mostly under 150,000 a month, except for the first five months. While the previous period accounted for over 150,000 most of the time.

On the other hand, the percentages of outside-the-spread crossings seemed to rise month after month. The proportions were constant at 1%–2% in the years 2013 and 2014, but then they increased significantly to 3%–8% monthly in the year 2015.

The numbers and proportions during this period demonstrate that the rule amendment effectively eliminated the 'at-the-spread' trades from the dark pool. However, they were still not 100% effective

Table 1: *The Number and Proportion of Internal Crossings Before 26 May 2013*

Year	Month	Within	% Within	At	% At	Outside	% Outside
2012	Jan	10,227	62.94	5,983	36.82	39	0.24
2012	Feb	45,017	37.24	75,048	62.08	820	0.68
2012	Mar	198,931	48.67	207,144	50.68	2,652	0.65
2012	Apr	146,084	40.59	211,595	58.80	2,203	0.61
2012	May	213,117	46.40	242,364	52.76	3,857	0.84
2012	Jun	247,315	52.04	223,157	46.96	4,760	1.00
2012	Jul	177,914	42.26	238,005	56.54	5,047	1.20
2012	Aug	230,979	47.63	247,916	51.13	6,015	1.24
2012	Sep	166,210	47.51	176,284	50.39	7,347	2.10
2012	Oct	80,996	36.26	140,390	62.86	1,966	0.88
2012	Nov	81,810	37.25	135,636	61.76	2,160	0.98
2012	Dec	84,527	36.60	144,647	62.63	1,763	0.76
2013	Jan	145,272	40.96	206,833	58.32	2,548	0.72
2013	Feb	183,441	41.89	249,579	56.99	4,917	1.12
2013	Mar	247,683	52.50	218,209	46.26	5,851	1.24
2013	Apr	321,794	59.15	215,301	39.58	6,920	1.27
2013	May	243,211	55.44	190,901	43.51	4,596	1.05

due to the proportion of `outside-the-spread` rising to 8% at the end of the year 2015.

Figure 1 exhibits the movement better. As we can see from the barplot, the proportion of `at-the-spread` crossings significantly decreased after May 2013. The proportion was then dominated by `within-the-spread` crossings, as the rule amendment only allowed meaningful price improvement on the dark pool. There were still some breaches of the TWPI rule, and the proportion was increasing towards the end of 2015. However, the number and proportion were not significant compared to the compliance of dark trades.

Economic Implication

Comparing the dollar value of the trades can be useful to examine the economic implication of the rule change. We compare 100 days before and after the rule change on 26 May 2013. The reason for using a smaller period of time is to control the change in trend of transactions as well as minimize the impact of macro- and microeconomic change. One hundred trading days before the TWPI rule amendment, `at-the-spread` crossings had a value of $3.97 billion, while 100 trading days later the value dropped to just $216 million. This decrease in value indicates that there was a movement of around $3.7 billion from `dark pool` to ‘lit pool’ from the stocks listed on the ASX200 index alone.

This number was quite high, given that this was only from the stocks listed on the ASX200 index and only accounts for 100 trading days. With this number alone, ASX could charge around $45,000 based on the trading fee on `lit market` around 0.15 basis point. The number will rise much higher if we consider the whole transaction on all stocks listed on ASX.

Table 2: *The Number and Proportion of Internal Crossings After 26 May 2013*

Year	Month	Within	% Within	At	% At	Outside	% Outside
2013	May	83,723	96.70	1,774	2.05	1,080	1.25
2013	Jun	246,549	95.92	6,982	2.72	3,506	1.36
2013	Jul	170,658	95.58	4,194	2.35	3,705	2.07
2013	Aug	193,579	97.10	3,324	1.67	2,449	1.23
2013	Sep	190,732	97.03	3,571	1.82	2,274	1.16
2013	Oct	171,878	97.81	2,279	1.30	1,570	0.89
2013	Nov	122,955	97.87	1,477	1.18	1,193	0.95
2013	Dec	146,428	97.43	2,175	1.45	1,691	1.13
2014	Jan	134,729	97.71	1,460	1.06	1,701	1.23
2014	Feb	135,040	97.69	1,528	1.11	1,668	1.21
2014	Mar	137,176	97.81	1,233	0.88	1,839	1.31
2014	Apr	99,761	98.00	840	0.83	1,200	1.18
2014	May	92,872	98.06	812	0.86	1,030	1.09
2014	Jun	114,112	97.51	1,023	0.87	1,887	1.61
2014	Jul	108,816	97.45	892	0.80	1,950	1.75
2014	Aug	101,156	96.76	774	0.74	2,613	2.50
2014	Sep	112,515	97.11	743	0.64	2,600	2.24
2014	Oct	104,250	96.85	1,135	1.05	2,258	2.10
2014	Nov	187,119	98.01	1,161	0.61	2,633	1.38
2014	Dec	78,916	94.52	722	0.86	3,851	4.61
2015	Jan	106,730	95.68	1,351	1.21	3,470	3.11
2015	Feb	78,793	94.61	705	0.85	3,788	4.55
2015	Mar	98,910	95.02	890	0.85	4,296	4.13
2015	Apr	74,511	95.34	640	0.82	2,999	3.84
2015	May	92,631	93.85	879	0.89	5,195	5.26
2015	Jun	94,497	93.99	1,054	1.05	4,989	4.96
2015	Jul	104,670	93.91	1,043	0.94	5,740	5.15
2015	Aug	106,823	91.68	1,030	0.88	8,669	7.44
2015	Sep	152,337	90.44	1,470	0.87	14,637	8.69
2015	Oct	155,282	91.48	1,664	0.98	12,794	7.54
2015	Nov	118,938	90.23	1,292	0.98	11,587	8.79
2015	Dec	102,508	85.96	1,581	1.33	15,156	12.71

Table 3: *The Percentage of Within/At/Outside-the-Spread Internal Crossings Before and After the Rule Change*

Marks	Prop_Before	Prop_After
Within	46.94	95.54
At Spread	52.00	1.23
Outside	1.05	3.23

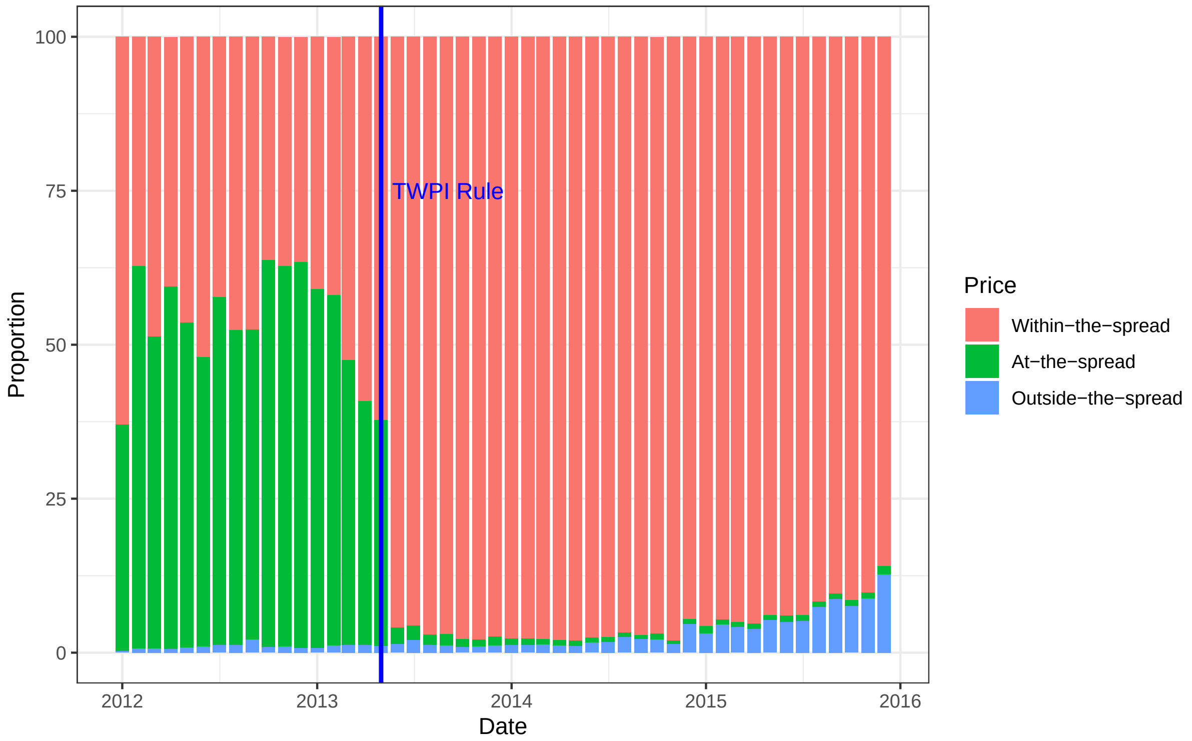

Figure 1: *Proportion of each trade price groups from Januari 2012 until December 2015. Red dash line is when the meaningful price improvement rule introduced.*

From the report by Charles Lane Advisory, *Review of recent rule changes affecting dark liquidity* (2014), it is known that there was a decline of total dollar volume in the below block size dark pool 100 days before and after the rule change, from \$54.9 billion to \$21 billion, meaning that there was a movement of more than \$33 billion from `dark pool` to `lit pool`. However, these numbers have taken into account all stocks in ASX, and the priority crossing mechanism occurred on the ASX trading platform, which we will not consider in this study. Based on these numbers, the `at-the-spread` crossings occurred on the ASX200 index-listed companies accounts only approximately 12% of that from all stocks listed on the ASX.

4.2 Violation Pattern After the Rule Change

Figure 1 shows that there were still violations of the TWPI rule after the rule change on 26 May 2013. If we look at the proportion, the percentages were going higher year after year. This trend is primarily due to the increasing trend of 'outside-the-spread' crossings, which accounted for more than 12% of all internal crossings in December 2015. These numbers could potentially rise in the years following 2015. However, this is out of scope of this study.

Since we have the broker IDs variable on our dataset, we can examine what kind of brokers tend to violate the TWPI rule. We classify brokers's types as institutional, retail, and HFT-proprietary. Figure 2 exhibits the top 10 brokers that have done the most violations. The plot reveals that only two brokers dominated the number of violations. Given that both brokers are institutional, could it be inferred that institutional brokers tend to commit more violations than other types of brokers? Since there were only two brokers dominating the violation, it is probably more like an individuality reason than because of the broker's type.

We also understand that regulators did not impose any penalties for these violations, given that the first release regarding the TWPI rule violation occurred in February 2023. This is probably because the rule has just been introduced and the number of violations was still insignificant compared to all internal crossings. However, as the proportion was on a rising trend, regulators should have taken action to prevent the number from getting more significant, especially since it has already happened since 2015.

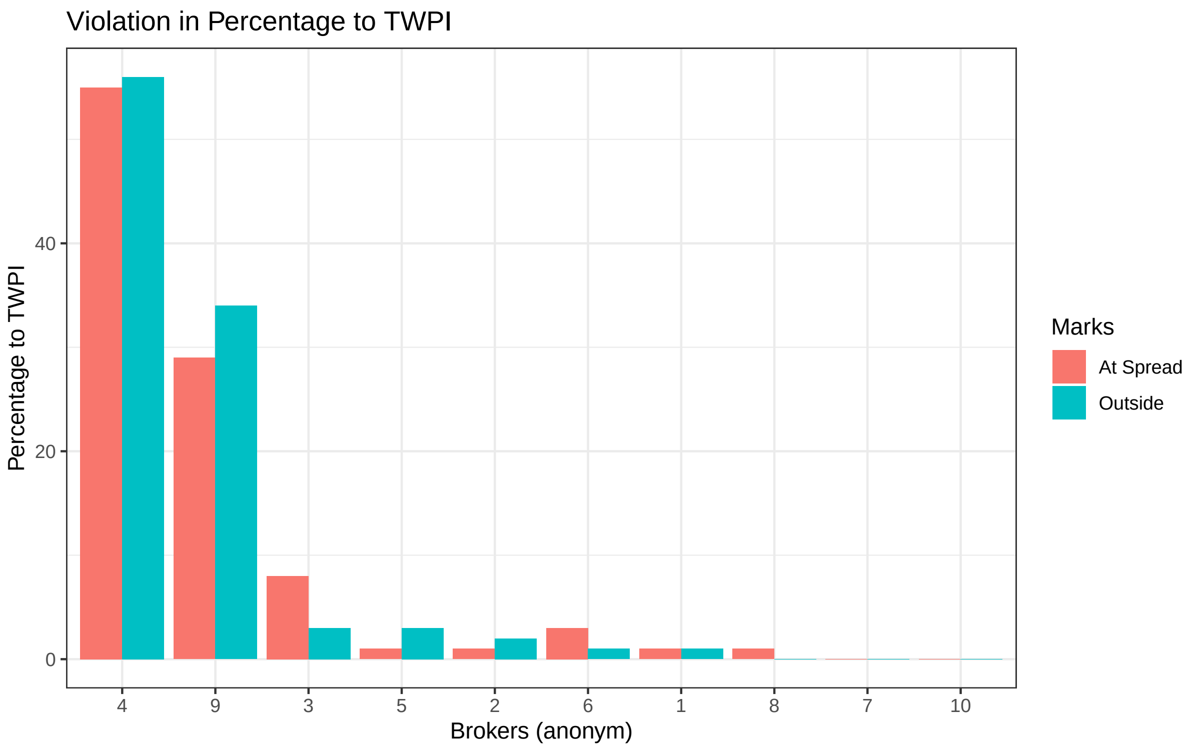

Figure 2

Modelling

To understand what condition contributed to the occurrence of a violation of the TWPI rule, we will do simple modeling using logistic regression and random forest. Given that only two brokers dominated the violation, we plan to train the model using data from these brokers. For this reason, we did not include the broker's type as one of the predictor variables. If we put the broker's type as one of the predictors, then the model will take it as the dominant variable causing the violation.

To model the violation pattern, we use several predictor variables from the transaction data, namely dollar value, trade volume, hour, and date, and also market data, such as market spread and best

Table 4: *Confusion Matrix of the Test Set for Logistic Regression Model on `At-The-Spread` Violations*

At-The-Spread Violation			
Class	cl_acc	0	1
0	0.7043337	732681	307566
1	0.6811604	4539	9697

bid/ask volume. We also make models for each violation type, `at-the-spread` and `outside-the-spread` violations.

We split the data to train and test datasets. Moreover, we will check the robustness of the model using data from brokers other than these top two brokers. One thing that is important to consider is that this data has a very imbalanced response variable, as the violation only accounts for approximately 5% of all internal crossings reported as TWPI (see again Table 3).

Logistic Regression

Since our case is a classification problem, logistic regression is one of the most popular methods for predictive modeling. This model uses logistic function and maximum likelihood method to estimate the coefficients of each variable (James et al., 2021). The logistic function predicts the probability of belonging to one of two classes, with lower balance generally classified to class 0 and higher balance to class 1. A probability threshold can be set to determine the cutoff for classifying an observation into either class. In our case, observations with lower balances may fall into class "non-violation", while higher balances may fall into class "violation", depending on the chosen threshold. We use `roc()` and `coords()` functions from the `pROC` package to find the best threshold for given cases.

At-The-Spread Violation

After we train the model on the train dataset, we apply the fitted model to the test dataset to examine the accuracy of the `at-the-spread` violation model. Table 4 exhibits the confusion matrix for the result on this dataset. The class 0, meaning the non-violation trades, were correctly predicted as non-violation 732,681 times, while wrongly predicted 307,566 times. On the other hand, class 1, meaning the violation trades, were correctly predicted as violations 9,697 times, while wrongly predicted 4,539 times.

These numbers produce evaluation performance of sensitivity as high as 0.7043 and specificity as high as 0.6811. The balanced accuracy for this model on the test dataset is 0.6927. These numbers are quite high, given the complexity of the data and how imbalanced the response variables are.

Examining the significance of the variable from Table 5, we understand that most of the variables are significant predictors as the p-values are close to zero. The only variable that does not significantly contribute to the model is the trading hour. Dollar value and the market data, such as the spread and

Table 5: *Variables significance of the Logistic Regression Model (At-the-Spread Violation)*

term	estimate	p.value
(Intercept)	-11.3937554	0.6291550
DollarValue	-1.6006516	0.0000000
AskVol_before	-0.5139181	0.0000000
BidVol_before	-0.2499091	0.0000000
Volume	0.9221675	0.0000000
spread	-0.9670122	0.0000000
Hour10	7.0561430	0.7648868
Hour11	6.7464728	0.7749205
Hour12	6.5977296	0.7797535
Hour13	6.4605623	0.7842180
Hour14	6.4334577	0.7851010
Hour15	6.4081678	0.7859251
Date	0.0557625	0.0000000
Dayofweek3	0.1503556	0.0000000
Dayofweek4	0.0895771	0.0000072
Dayofweek5	0.1340555	0.0000000
Dayofweek6	0.0768478	0.0002401

Table 6: *Confusion Matrix of the Test Set for Logistic Regression Model on `Outside-The-Spread` Violations*

Outside-The-Spread Violation			
Class	cl_acc	0	1
0	0.6059501	614494	399606
1	0.7386276	10555	29828

best bid/ask volume, have a negative relationship with the violation, while the trade volume has a positive correlation. These mean that the violation might happen more on the condition where the market has a narrow spread but less volume depth.

Outside-the-Spread Violation

After we train the logistic regression model to understand the variables that contribute to `at-the-spread` violation, we then want to understand regarding `outside-the-spread` violation. Table 6 exhibits the confusion matrix for this violation modeling on the test set. The class 0, meaning the non-violation trades, were correctly predicted as non-violation 614,494 times, while wrongly predicted 399,606 times. Conversely, we correctly predicted class 1 as a violation 29,828 times and incorrectly predicted it 10,555 times.

These numbers yield evaluation performance as high as 0.6059 and specificity as high as 0.7386. The balanced accuracy for this model on the test dataset is 0.6723. These numbers are similar to the previous prediction model on the `at-the-spread` violation.

Differently, only the market depth variables have significant value on this logistic regression model:

Table 7: *Variables significance of the Logistic Regression Model (Outside-the-Spread Violation)*

term	estimate	p.value
(Intercept)	-11.3161525	0.6209963
DollarValue	0.0003271	0.9866395
AskVol_before	-0.6405505	0.0000000
BidVol_before	-0.4560265	0.0000000
Volume	0.0062374	0.7478965
spread	-0.6439136	0.0000000
Hour10	7.9783291	0.7273906
Hour11	7.9440813	0.7285145
Hour12	7.7628845	0.7344702
Hour13	7.6479714	0.7382556
Hour14	7.6383343	0.7385733
Hour15	7.4675265	0.7442124
Date	-0.0024128	0.4942947
Dayofweek3	0.0418110	0.0004022
Dayofweek4	-0.0111388	0.3406180
Dayofweek5	-0.0406654	0.0008428
Dayofweek6	-0.0331316	0.0070492

the best bid/ask volume and the market price spread. All three have a negative correlation with the `outside-the-spread` violation, hence consistent with previous modeling: violation might happen more on the condition where the market has a narrow spread but less volume depth.

Random Forest

Random forest is another popular method for classification modeling. This model is an ensemble of decision trees using bootstrapped training samples. Each tree is trained on different subsets of the data using a random sample of predictors (not all variables used in each tree). This method forces each split to consider the non-dominant variables (James et al., 2021). The random forest model predicts the probability of classes by aggregating the predictions from each individual tree, typically with the majority vote.

Unlike logistic regression, random forest does not assume a linear relationship between predictors and the outcome (James et al., 2021), making it well-suited to capture complex interactions between variables. We will use the `step_downsample` function from the `themis` package to handle the imbalance observation in the response variable.

At-The-Spread Violation

We fit the random forest model to the train dataset, then apply the fitted model to the test dataset. Table 8 shows that the class 0 (non-violation trades) were correctly predicted 825,000 times while wrongly predicted 215,240 times. On the other hand, class 1 was correctly predicted as a violation

Table 8: *Confusion Matrix of the Test Set for Random Forest Model on* `At-The-Spread` *Violations*

At-The-Spread Violation			
Class	cl_acc	0	1
0	0.7930886	825008	215239
1	0.8194015	2571	11665

Table 9: *Variables importance of the Random Forest Model (At-the-Spread Violation)*

	0	1	MeanDecreaseAccuracy	MeanDecreaseGini
DollarValue	59.007187	100.40460	131.27640	2617.498
AskVol_before	146.211349	287.26167	393.89538	4729.450
BidVol_before	142.931942	286.48835	364.34268	4790.141
Volume	49.769757	103.75083	134.10499	2248.948
spread	230.996712	431.11889	589.67343	8098.680
Date	12.170909	65.69846	59.37895	2057.638
Hour	40.580342	51.03897	66.12206	1264.557
Dayofweek	8.310214	49.13468	42.56112	1139.813

11,665 times while wrongly predicted 2,571 times.

This model has 0.7930 sensitivity and 0.8194 specificity. The balanced accuracy for this model on the test dataset is 0.8062. The accuracy measures are significantly higher than those of the previous model that used logistic regression.

Similar to the logistic regression model, we can examine which variables have most contributed to the model's accuracy. From variable importance in Table 9, we understand that the market spread and best bid/ask volume are the most important to the model, both based on the two metrics, `Mean Decrease in Accuracy` and `Mean Decrease in Gini`. This result supports the previous model where the best bid/ask volume and the market price spread are the most significant variables in the model.

Outside-The-Spread Violation

We then fit the random forest model to understand the `outside-the-spread` violation pattern. Table 10 shows that the class 0 (non-violation trades) were correctly predicted 851,567 times while wrongly predicted 162,533 times. On the other hand, class 1 was correctly predicted as a violation 33,640 times while wrongly predicted 6.743 times.

This model has 0.8330 specificity and 0.8397 sensitivity. The balanced accuracy for this model on the test dataset is 0.8364. A random forest model also predicts this type of violation better than logistic regression.

When examining the variable importance in Table 11, the model places the most importance on the variables from the market data, specifically the market price spread and best bid/ask volume. This

Table 10: *Confusion Matrix of the Test Set for Random Forest Model on `Outside-The-Spread` Violations*

At-The-Spread Violation			
Class	cl_acc	0	1
0	0.8397900	851631	162469
1	0.8331724	6737	33646

Table 11: *Variables importance of the Random Forest Model (Outside-the-Spread Violation)*

	0	1	MeanDecreaseAccuracy	MeanDecreaseGini
DollarValue	125.50860	157.4889	217.4215	6946.841
AskVol_before	204.88930	413.4406	504.1273	15512.206
BidVol_before	242.37856	432.8081	517.8487	15926.278
Volume	133.80397	150.5007	204.1081	6561.616
spread	171.97334	601.5576	720.5479	16535.967
Date	51.70494	261.0853	258.5334	7005.261
Hour	85.76693	130.0931	144.6198	3594.039
Dayofweek	32.07256	241.1619	236.0054	3561.868

pattern is similar to previous predictions on `at-the-spread` violations.

Robustness Check

To check how robust are the models we have on previous sections, we fit them on the whole new dataset: the internal crossings data from out of the top two brokers with the most violations. We present the accuracy metrics on Table 12. From the table, we can imply that the models are robust, as the accuracy metrics for this new dataset produce similar accuracy. These metrics clearly demonstrate the superiority of random forest models over logistic regression in this scenario, with approximately 10% more balanced accuracy for each type of violation.

4.3 Limitation

These are the limitations of this study that may impact the accuracy of the result:

- **Data Imbalance**: Only about 5% of all internal crossings are violations, hence the dataset is extremely imbalanced. This condition may hinder the models' ability to accurately predict the violation pattern.

Table 12: *Accuracy Metrics to Measure Models Robustness - Using Dataset from Other Than Top 2 Brokers*

Violation	Model	Sensitivity	Specificity	Bal_Accuracy
At-the-Spread	Logistic Regression	0.7013346	0.6390095	0.6702
At-the-Spread	Random Forest	0.7642962	0.7763173	0.7703
Outside-the-Spread	Logistic Regression	0.5825263	0.7178747	0.6502
Outside-the-Spread	Random Forest	0.8387124	0.6734735	0.7559

- **ASX200 Index Stocks**: This study uses only stocks from the ASX200 index due to the limited processing power. Therefore, trends that are unique to non-ASX200 stocks might be missed if they are left out.
- **Not the most recent data**: SIRCA provides ASX transaction data only up to 2016. Hence, this study does not take into consideration data with the most recent trading patterns and behavior.

5 Conclusion

When the TWPI rule went into effect on 26 May 2013, it effectively reduced the `at-the-spread` crossings in the `dark pool`, moving them to the `lit market`. Approximately $3.7 billion worth of trades went from the dark pool to the lit market in the 100 trading days after the rule change. This shows that the rule had a big effect on how trades were distributed.

In addition, violations are still insignificant, with the proportion of around 5% of all internal crossings. However, monthly numbers show that violations are going up, and by December 2015, they had reached 12%. Looking at the pattern, there are a lot of violations involving just two brokers, which makes it less reliable to attribute violations to certain kinds of brokers. After applying machine learning modeling to the data, we understand that the most important factors for both types of violations (`at-the-spread` and `outside-the-spread`) were the market conditions, especially the price spread and market depth, which could be seen by the best bid/ask volume.

6 Potential Future Work

- Add Variables in Violation Analysis. To gain a deeper understanding of violation patterns, future research should include more variables including market capitalization, industrial sector, volatility, and lagged price values.
- Using Other Analysis Techniques. Other approaches, such as factor modeling or Principal Component Analysis (PCA), may give more insight on the directional influence of variables.
- Extend Research Questions.
  - Evaluate the financial effects from the breach of the TWPI rule.
  - Identify the reasons why market spread and market depth correlate with the violation pattern.

7 References

ASIC Market Integrity Rules (Competition in Exchange Markets) 2011.

Australian Securities and Investments Commission. (20 May 2004). *Market demutualisation and privatisation: The Australian experience* [Speech]. Retrieved 24 October 2024, from https://asic.gov.au/about-asic/news-centre/speeches/market-demutualisation-and-privatisation-the-australian-experience/

Australian Securities and Investments Commission. (May 2014). *Report 394: Review of recent rule changes affecting dark liquidity* (PDF). Retrieved 24 October 2024, from https://download.asic.gov.au/media/1344596/rep394-published-19-May-2014.pdf

Australian Securities and Investments Commission. (May 2022). *Market integrity update: Issue 137*. Retrieved 24 October 2024, from https://asic.gov.au/about-asic/corporate-publications/newsletters/market-integrity-update/miu-issue-137-may-2022/

Australian Securities and Investments Commission. (3 February 2023). *Wilson's Advisory and Stockbroking Ltd ACN 010 529 665 pays $548,328 infringement notice* [Media release]. Retrieved 24 October 2024, from https://asic.gov.au/about-asic/news-centre/find-a-media-release/2023-releases/23-016mr-wilsons-advisory-and-stockbroking-ltd-acn-010-529-665-pays-548-328-infringement-notice/

Australian Securities Exchange. (n.d.). ASX regulatory framework. Retrieved 24 October 2024, from https://www.asx.com.au/about/regulation

Australian Securities Exchange. (n.d.). ASX story. Retrieved 24 October 2024, from https://www.asx.com.au/about/asx-story

Australian Securities Exchange. (n.d.). About Us. Retrieved 24 October 2024, from https://www.asx.com.au/about

Australian Securities Exchange. (19 February 2024). *ASX Trade: Markets Participant and Trading Schedule of Fees* (PDF). Retrieved 24 October 2024, from https://asxonline.com/content/dam/asxonline/public/documents/schedule-of-fees/asx-trade-markets-participant-and-trading-schedule-of-fees.pdf

Charles Lane Advisory Pty Ltd. (May 2014). *Report by Charles Lane Advisory: Review of recent rule changes affecting dark liquidity* (PDF). Attachment to ASIC Report 394: Review of recent rule changes affecting dark liquidity. Retrieved 24 October 2024, from https://download.asic.gov.au/media/1344590/rep394-attachment-published-19-May-2014.pdf

Duong, H. N., Lajbcygier, P., Lu, J. S., & Vu, V. H. (2018). The effect of anonymity on price efficiency: Evidence from the removal of broker identities. *Pacific-Basin Finance Journal, 51*, 95–107. https://doi.org/10.1016/j.pacfin.2018.06.004

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2nd ed.). Springer.

Securities Industry Research Centre of Asia-Pacific. (n.d.). *About SIRCA*. SIRCA. Retrieved 26 October 2024, from https://www.sirca.org.au/about-sirca/