Priority Rules [1]

Hans Degryse [2] and Nikolaos Karagiannis [3]

This version: October 2020

[1]We thank Michael Brolley, Sabrina Buti, Fabio Castiglionesi, Jean-Edouard Colliard, Carole Comerton-Forde, Sarah Draus, Ariadna Dumitrescu, Sean Foley, Manuela Geranio, Carole Gresse, Björn Hagströmer, Frank M. Hatheway, Terrence Hendershott, Peter Hoffmann, Johan Hombert, Vincent van Kervel, Michael Koetter, Evangelos Litos, Andrew W. Lo, Mike Mariathasan, Jose Mendoza, Albert J. Menkveld, Dagfinn Rime, Tarik Roukny, Batchimeg Sambalaibat, Norman Schürhoff, Duane Seppi, Gunther Wuyts, Bart Z. Yueshen, Darya Yuferova, Marius A. Zoican and seminar participants at Banca d'Italia, KU Leuven, the Luxembourg School of Finance, the Tinbergen Institute Amsterdam, Universita Cattolica Milano, UPF, the Group of Economic Advisors meeting of ESMA, the CEPR 'I have seen the future' conference in London, the 2018 CEPR European Summer Symposium in Financial Markets (main program), the 2018 CEPR-Imperial-Plato Market Innovator Conference, the 2^{nd} SAFE Market Microstructure Conference, the 35^{th} Annual Conference of the French Finance Association, the 2^{nd} European Capital Markets Workshop, the IFABS 2018 Conference, the 3L workshop, the 3^{rd} CEPR Annual Spring Symposium in Financial Economics, the SGF Conference, the 9^{th} Erasmus Liquidity Conference, the 3^{rd} Dauphine market microstructure workshop, the 3^{rd} EBC Network Workshop, and the 11^{th} Swiss Winter Conference on Financial Intermediation for useful comments. Financial assistance by the French National Research Agency (ANR) through project GHOST and the KU Leuven (OT grant) is gratefully acknowledged. A previous version of the paper circulated as 'Once Upon a Broker Time? Order Preferencing and Market Quality'.

[2]KU Leuven and CEPR. Faculty of Economics and Business. Department of Accounting, Finance and Insurance, Naamsestraat 69, 3000 Leuven, Belgium. e-mail: hans.degryse@kuleuven.be

[3]KU Leuven. Faculty of Economics and Business. Department of Accounting, Finance and Insurance, Naamsestraat 69, 3000 Leuven, Belgium. e-mail: nikolaos.karagiannis@kuleuven.be

Abstract

While regulators often mandate price priority across markets, they do not impose secondary priority rules. Order preferencing by a broker to a specific market may then serve as tiebreaker. We compare order preferencing, modeled as price-broker-time priority (PBT), to price-time priority (PT). The secondary priority rule determines a limit order's execution probability, and hence investors' choice between limit and market orders. When the tick is tight relative to the dispersion in investors' valuations, trading rates are higher with PBT whereas investor welfare is higher with PT. The opposite holds for wide ticks. Our model has empirical and regulatory implications regarding market fragmentation. *JEL* Code: G10.

1 Introduction

In most countries, regulation mandates price priority when trading financial assets in fragmented markets. For example, the Securities and Exchange Commission (SEC) rule 611 of Regulation National Market System (also know as the 'Order Protection Rule' or the 'Trade-through Rule') implements price priority among venues. Even though best-execution standards require price priority across venues, they do not request time priority. When multiple venues display the best price, a broker can route customer orders to any of those platforms. A natural question that arises: should time priority be enforced *across* and *within* platforms, or are other secondary priority rules preferable? Is there a "one size fits all" priority rule, or should priority rules be adjusted according to the underlying trading needs? In this paper, we address these questions by studying how priority rules affect investors' mode of market participation (market versus limit orders), investor welfare, and the way how markets fragment. We further study which priority rule is socially preferred and whether regulatory intervention is required.

Regulation has fostered fragmentation of trading (e.g., Regulation National Market System (RegNMS) in the U.S. and the Markets in Financial Instruments Directive (MiFID) in Europe). While the 'trade-through prohibition' of RegNMS enables investors to access the best price for a security, time priority is broken when trading fragments *across* trading venues. A broker's handling of retail-client orders then determines to which venue an order is directed.

IOSCO (2017) identifies several incentives that could direct a broker's orders to a particular venue. First, monetary benefits received from trading venues in the form of favorable trading fees may determine where a broker sends its retail-client orders (e.g., Colliard and Foucault, 2012; Foucault, Kadan and Kandell, 2013). Battalio, Corwin and Jennings (2016) note that fees and rebates to brokers are not generally passed on to retail clients. Volume-based fees also could encourage a broker to concentrate its trading on one venue (e.g., Spatt, 2019).[1] Second, a broker may be an important shareholder of an

[1]The SEC scrutinizes the use of tier-based pricing, see https://www.reuters.com/article/us-sec-exchanges-fees-exclusive/exclusive-u-s-sec-scrutinizes-fairness-of-stock-exchange-pricing-idUSKCN1QO2CY

exchange, and thus have an affiliated venue. A broker then has incentives to route orders to that broker-affiliated venue (e.g., Anand et al., 2019). In this paper, we compare a setting where price-time priority is enforced across trading venues to a setting where each broker has a preference for one venue as secondary priority rule.

Priority rules not only play a role *across* venues but also *within* a venue. Each U.S. trading venue currently manages its limit order book according to price-time priority (PT). PT implies that the first order at a new price becomes the first to trade at that price, and any subsequent orders are executed in the time order in which they are received. Other trading venues employ different secondary rules than time. Some venues impose price-broker-time priority (PBT). With PBT, time priority is violated as orders stemming from the same broker will execute against each other even if that broker's order was not the first in the queue. Examples of trading venues with PBT are those in Canadian markets (e.g., Toronto Stock Exchange), the Nordic countries (e.g., NASDAQ OMX), and continental Europe (e.g., Euronext's Internal Matching Service). In the U.S., Investors Exchange (IEX) had PBT while being an alternative trading system but when becoming a national securities exchange in September 2016, its priority rule became PT. In this paper we compare market quality and investor welfare across price-time and price-broker-time priority settings, and study whether brokers or venues have incentives to adopt PBT or prefer PT.[2]

Other examples of violation of price-time priority involve displayed versus undisplayed orders, and off-exchange trading. Most venues have a price-display-time priority implying that hidden orders effectively are at the back of the queue, and that iceberg orders lose time priority on their undisclosed part (e.g., Buti and Rindi, 2013). Off-exchange reported trades do not obey price-time priority. And Blockchain could also impact the priority structure of trading and post-trading as miners confirming trades exhibit different abilities in doing so, or ask different fees according to the speed in which

[2]PT was not always the only allocation rule in U.S. financial markets. In 1996, while certain U.S. exchanges were still allowed to offer PBT, U.S. Congress had the SEC conduct a study on the effects of the practice. The SEC's "Report on the Practice of Preferencing" found no proof that it had negative effects on the market, but added that the "findings should not be taken to mean that the Commission believes that such adverse effects may not arise in the future". IEX's decision to become a national securities exchange coincided with a switch from PBT to PT.

trade confirmation is required.

Our model is build to capture the essentials of priority rules *within* or *across* markets. An investor employs one out of two brokers. Investors submit market orders (MOs) or limit orders (LOs) through their broker, or abstain from trading. In a *single market* interpretation, LOs are stored in one limit order book (LOB). PT implies that the arrival time of the investor's LO is the secondary priority rule. A MO then executes against the first submitted LO at that price. In contrast, PBT has own broker as secondary priority rule such that an arriving MO executes against LOs submitted by investors using the same broker when at the same price and whenever possible. In the absence of investor LOs from any broker, MOs execute against dealer-specialists who then supply liquidity.[3]

In the *multiple markets* interpretation, two broker-affiliated platforms coexist. Every broker puts its investors' LOs in the LOB of its own broker-affiliated platform. The priority rules in place then determine how MOs are allocated across the two platforms. With PBT, MOs execute against LOs submitted by investors having as secondary priority rule a preference for the own-broker-affiliated platform; in the absence of a LO in both books, they execute against the dealer-specialist. Markets are then *fragmented*. With PT, and when the same price prevails in both LOBs, MOs execute against an investor's LO that arrived first in either one of the LOBs. PT then integrates back the two platforms and thus leads to a *centralized* market.

To gain focus, we concentrate on the single market interpretation. However, we also stress insights that are important and unique to the multiple markets interpretation (see Section 7). To do so, we build upon the work of Parlour (1998), Foucault (1999), Colliard and Foucault (2012), or Hoffmann (2014), and model a one-tick limit order book with infinite horizon where traders with different valuations for an asset arrive sequentially. We extend these models by allowing LOs to stay in the LOB for two periods. This requirement is the minimum needed to perform a meaningful study of the impact of PBT (i.e., allowing that limit orders can jump the queue when from the same

[3]This is in line with a hybrid LOB (e.g., NYSE) where designated market makers or specialists provide liquidity in the LOB and the exchange regulation requires them to yield both price and time priority to investors' orders (e.g., NYSE rule 72 required such priority; for further discussion, see de Jong and Rindi, 2009, or Foucault, Pagano and Roell, 2013).

brokers), and at the same time keep the model tractable. In our main analysis and for tractability, we incorporate dealer-specialists that provide liquidity at the minimum tick in the absence of LOs submitted by investors. Arriving traders then can always submit a MO independent of previous traders having put a LO in the book. This restricts the number of relevant states influencing investors decisions. We identify the stationary probability distribution of the system, and compute trading rates and investor welfare per period.

Our analysis generates novel insights about market quality (depth, fill rates of LOs, and trading rates) and investor welfare. First, priority rules determine overall trading rates, the degree of dealer-specialist participation, fill rates of LOs, and investor welfare. Two different forces drive the results when comparing PBT to PT: (i) the *anticipation effect* stemming from the threat of being queue-jumped makes a MO more attractive to first-in-line investors (i.e., arriving investors not facing a competing LO at the time of arrival), resulting in a higher trading rate under PBT; (ii) investors have greater incentives to join the queue under PBT, exploiting the possibility of preferential execution (i.e., the *queue-joining effect*). The importance of these two forces hinges on the size of the minimum tick relative to the dispersion in traders' valuations (for short – minimum tick or tick). When the minimum tick is small, trading rates are higher with PBT compared to PT. The anticipation effect then dominates the queue-joining effect as the spread is low. The anticipation effect also leads to LOBs that are more likely to be empty, and to trades where investor and dealer-specialist are each other's counterparty (i.e., customer-dealer/specialist trades). When the minimum tick is wide, the queue-joining effect dominates the anticipation effect as investors are more likely to participate in the market. Trading rates are higher with PT than PBT, but customer-customer trades are higher with PBT. Independent of tick size, fill rates of LOs are higher under PBT. Investor welfare is higher with PT than with PBT when the tick is small whereas the opposite holds for large ticks. With small (large) ticks, the composition of trades is less (more) favorable to generate investor welfare with PBT.

Second, priority rules determine the depth of the LOB and thus the aggregate depth of the market. With PT, LOBs have a higher average depth than with PBT. LOBs that are "shallow" (i.e., empty) and "deep" (i.e., depth of 2) are more prevalent with PBT compared to PT. The *anticipation effect* of being queue jumped as well as actual queue jumping creates more often an empty book. The *queue-joining effect* explains the result on "deep" books.

Third, PT and PBT generate different systematic patterns in trade and order flow. While LOs at one side of the market are more likely followed by LOs at the same side under PBT, the opposite applies for MOs. With PT, an investor's broker-affiliation does not impact order submission strategies. In contrast, with PBT, two consecutive LOs at the ask are less likely to come from same-broker investors than from different-broker investors.

Fourth, our model has implications for market design. When brokers are given the option to adopt PBT or PT, and assuming brokers maximize their investors' welfare, PBT results as an equilibrium outcome. When the minimum tick is small, brokers are in a prisonner's dilemma situation. While both jointly would be better off with PT, it is a dominant strategy to offer PBT. The market outcome then differs from the socially preferred one. For wide ticks, the market outcome and socially preferred outcome coincide as both yield PBT.

Fifth, when considering the *multiple markets* setting, we show that even when the two platforms are identical in terms of affiliated traders, order flow on the two platforms is dependent upon each other with PBT, i.e., $fragmented$ markets. In particular, limit orders on one platform are more likely to be followed by limit orders on the other platform than on the same platform.

Finally, we discuss several extensions including opacity about a standing LO's broker affiliation, more than two brokers, the absence of dealer-specialists, dealer-specialists quoting at inferior prices, and other priority rules (random matching). While obviously these have some impacts as they affect the magnitude of the anticipation and queue-joining effects, the main takeaways from our analysis remain.

Viewing PBT as a "weak" version of a pro-rata allocation (for more details, see, Section 8.6), many of our model's theoretical outcomes and empirical implications are supported by the limited empirical research in trading allocation rules. Lepone and Yang (2015), exploit the introduction of a pro-rata allocation to Euribor futures contracts on NYSE LIFFE and find that not only trading volume but also trading frequency (trading rates) increase substantially after the event for the "event contracts", similar to what our model predicts.[4] Further, Aspris et al., (2007) and Haynes and Onur, (2019), investigate the transition from a pro-rata allocation rule to a time pro-rata (first-in-first out) trading protocol.[5] Aspris et al., (2007) verify that after the repeal of the pro-rata, the (value) depth for the futures contracts decreases, while Haynes and Onur (2019), show that the volatility in the order book depth is higher under a pro-rata matching algorithm, both consistent to the empirical implications of our model.[6]

Our paper contributes to several strands of literature. First, and probably most closely related to our work are the papers on multiple markets by Foucault and Menkveld (2008) and van Kervel (2015). These papers model how liquidity providers determine the length of the queue in each market, anticipating liquidity demanders to trade a number of shares drawn from a probability distribution. Liquidity providers have access to both markets whereas some liquidity demanders may only access one market. Foucault and Menkveld (2008) show that the joint depth of all books increases relative to the case where only one book is present. van Kervel (2015) additionally shows how liquidity providers modify their orders after liquidity demanders submit trades that may contain information. These papers show that the possibility of queue jumping increases the aggregate depth of markets. In contrast, our model allows traders to chose between limit orders and market orders. We find that secondary priority rules influence this choice such that queue jumping stemming from multiple platforms may lead to less aggregate depth provided by investors.

Second, we extend limit order book models such as Parlour (1998), Foucault (1999), Handa, Schwartz and Tiwari (2003), or Rosu (2009). These models assume PT and

[4]The Euribor futures contracts constitute a tight market.

[5]Aspris et al., (2007) investigate three futures contracts traded in the NYSE LIFFE, while Haynes and Onur, (2019), study the 2-year Treasury futures traded in the Chicago Mercantile Exchange.

[6]In our model, depth and "value" depth coincide, since traders transact only one unit of asset.

study patterns in order and trade flow. We incorporate an additional priority layer in the trade allocation rule, and allow LOs to stay in the book for two periods.

Third, our paper relates to recent work modeling over-the-counter markets featuring trading via marketmakers (Duffie et al., 2005), bilateral bargaining (Duffie et al., 2007), or a LOB with random matching (Dugast, 2018). Order flow in these setups stems from traders switching between "high" or "low" preference for asset ownership. Similar to these models, we focus on equilibria in which aggregate preferences are in a steady state limiting the dimensionality of the state space. As in Dugast (2018), our setup has a spread equal to the minimum tick, and LOs queue before executing or being canceled. We study how priority rules impact investors' choices between MOs and LOs, and feature investors' with a continuum of personal valuations for an asset.

Fourth, our paper links up with recent literature on queuing and speed in LOBs as well as on sub-penny trading (i.e., offering a meaningless price improvement to jump the queue in the LOB). Tick size creates rents for liquidity provision determining the length of the queue and the type of liquidity providers (Chao et al., 2017; Wang and Ye, 2017; Yueshen, 2014). We show that priority rules impact the length of the queue. Buti et al., (2015) investigate how sub-penny trading impacts the market quality and welfare on the public LOB. They find that sub-penny trading is higher when the public book has high liquidity or a high tick-to-price ratio. Sub-penny trading negatively impacts liquidity in the public book. We model the more general practice of PBT of which sub-penny trading is one example.

Fifth, our work relates to the literature on order preferencing as tie-braking rule and the driving forces for the multimarket trading. The SEC (1997) report mentions that "there are numerous practices by which a broker-dealer may obtain time priority over pre-existing customer orders." Past empirical work has found that preferencing could have negative effects on market quality (e.g., Bloomfield and O'Hara, 1998; Chung et al., 2004). Parlour and Seppi (2003) model intermarket competition with preferencing as a tie-braking rule when indifferent. They show that such preferencing for one or the other market substantially impacts the viability of particular market designs. Our paper focuses on priority rules *across* and *within* markets, and shows that tie-braking rules influence market outcomes. Our model explains why trading may fragment across multiple venues as brokers may endogenously select to trade on different venues. This provides an additional explanation on the increased number of trading platforms that we observe. The idea that queue jumping contributes on how markets fragment, is also present in recent empirical work (e.g., Foley et al., 2019).

Finally, our paper adds to the work on maker-taker pricing and routing decisions by brokers (e.g., Colliard and Foucault, 2012; Foucault, Kadan and Kandell, 2013 for make-take pricing; and Battalio, Corwin and Jennnings, 2016, and Cimon, 2018, for routing decisions). With equal prices, the "net price" may differ due to differences in take fees. Harris (2013) argues that this allows sophisticated traders to step in front of others. As retail traders typically delegate the routing decision to brokers, and brokers do not pass on make-take fees, Angel, Harris and Spatt (2011) argue this leads to a conflict of interest in the broker's routing decision (see also Malinova and Park, 2015). Our paper investigates such a situation where each broker prefers to route its client orders to a different venue.

The remainder of the paper is organized as follows. Section 2 presents the set up of the model. In Section 3, we analyze the consequences of PT and PBT to market quality and investor welfare. Section 4 studies the adoption of priority rules. Next, in Section 5 we examine the implications of priority rules to trade composition and to the number of dealer-specialists. Section 6 identifies the testable implications and provides regulatory insights derived from our model. Section 7 discusses predictions which are unique to our multiple markets interpretation. In Section 8, we discuss several extensions and Section 9 concludes the paper.

2 Model

2.1 General setup

We consider an infinite horizon discrete time model of Parlour (1998) describing the market as a one-tick limit order book (LOB), but add the presence of dealer-specialists. We derive investors' optimal decisions regardless the specific time period they arrive to the market, and thus identify steady state equilibria. Investors anticipate the rational behavior of future arriving investors, and similar investors take similar decisions independent of t when faced with the same state of the LOB.

Following Foucault (1999), Goettler et al., (2009), Colliard and Foucault (2012), and Hoffmann (2014), we model a market using discrete time intervals. Our market consists of one asset with value $Z_t = Z$, not subject to innovations. Z consists of two components; a common component Z_0 and a private component bV with b a parameter drawn from a uniform distribution β with support $[0, 2]$, i.e., $Z = Z_0 + bV$. This is common knowledge. At each time t, an investor arrives, willing to trade one unit of the asset. The investor is with equal probability a buyer or a seller. An investor comes only once to the market and has to decide at t among the following three choices. She may submit a market order (MO), place a limit order (LO) or refrain from trading. Once her decision is made it cannot be altered and a LO cannot be canceled, withdrawn or modified. Her LO is either filled within the next two periods or expires after that. LOs staying in the LOB for two periods is the minimum needed to make a meaningful comparison between PBT and PT.[7]

All agents are risk-neutral, and maximize their expected utility from trading. Since we solve for a steady state equilibrium in an infinite time horizon game, we drop the time subscripts for notational simplicity, unless we want to emphasize the time sequence of events. We denote the ask and bid price by A and B, respectively. For simplicity, the mid-quote $((A + B)/2)$ is identical to the fundamental value of the asset $Z_0 + V$.

[7]Investors thus have trading opportunities during two periods and do not have a discount rate within those periods. Adding a discount rate would further favor market orders. We abstain from introducing this additional feature to focus on the key aspect of our paper being priority rules. Exogenous order cancellation is a common assumption when modeling LOBs (e.g., Hoffmann, 2014)

In order to avoid meaningless price undercuts, and to identify the impact of secondary priority rules, following Parlour (1998) and Degryse et al., (2009), we assume that intense competition has set the ask and bid prices at the minimum tick Δ (i.e., $\Delta = A - B$).[8] We relax this assumption in Section 8.[9]

Every investor is affiliated to one out of two brokers (X and Y) with her affiliation being randomly assigned. Brokers are used to pass on orders to the market, and have equal market shares of buyers and sellers. We consider the case of more than two brokers in Section 8. In our model 'brokers' should be interpreted as secondary priority rule when multiple LOs are available at the same price. So brokers are our way of modeling order preferencing. We discuss the implications for the multimarket interpretation in Section 7.

Every investor has a private trade-off between immediate and future consumption. This trade-off is captured by the private valuation bV for the asset, where b is a parameter drawn from a uniform distribution β with support $[0, 2]$ (see, e.g., Buti et al., 2017).[10] Upon arrival, an investor can submit a MO, or opt for a better price by posting a LO, but face a non-execution risk. For example, a value of b closer to zero is more likely to lead a seller to submit a MO to sell, since she has almost no private valuation for the asset. The decision of the arriving agent is also influenced by the state s of the LOB. In our model, broker affiliation (when PBT applies) and the length of the competing queue are two key determinants for her decision, since these affect the execution probability of her limit order. The execution probability thus depends on the secondary priority rule in place.

The arriving investor at t maximizes her utility by solving a dynamic problem which depends on the behavior of future arriving investors, their type and private valuations. The investor's action depends on her inclination (buyer or seller), her information about

[8] The choice for a minimum tick is further supported by both theoretical and empirical literature (see, e.g., Dugast, 2018; O'Hara et al., 2019).

[9] For an exposition of one tick futures market, see, Field and Large (2008).

[10] Hendershott and Menkveld (2014) document a uniform distribution with private values spanning 370 bps. Decomposing the asset's fundamental value to a common to all traders part Z_0, normalized to 53.04 and to an idiosyncratic (trader specific) component, captured by the term bV, then our distributional assumptions become equivalent.

the state of the order book, her private valuation, and under PBT her broker affiliation.[11] An investor determines the value b that would make her indifferent between trading with a MO, a LO or refrain from trading. In the absence of standing LOs submitted by investors, we assume the presence of dealer-specialists who then supply liquidity. This simplification implies that an arriving investor can always submit a MO, either against a standing LO, or against the dealer-specialist (see, e.g., Parlour and Seppi, 2003). We consider dealer-specialists as having a private valuation of one for the asset and a cost to be around on the market.[12]

We set the stage by an example. Assume that the arriving trader is a seller, affiliated with broker X and faces a state s in the book. She has expectations about the future arrivals of traders given our assumptions. She solves her decision problem in which she determines a private valuation b_0 which would make her indifferent between a MO or a LO. Since she is a seller, if her private valuation b satisfies the inequality $b < b_0$, then she submits a MO to sell. Let $P = P(A, B, X, Y, s)$ be the probability of execution of a LO then her expected gains from a sell LO are $P(A - (Z_0 + bV))$. We can immediately see that for large values of b the seller will refrain from trading. In particular for any $b > (A - Z_0)/V$ she prefers not to trade. For any value of b, in the interval $[b_0, (A - Z_0)/V]$, the trader opts for a LO.

For the following, we denote by $b_k^S(s),\ k \in \{x, y\}$ the private valuation of a seller, affiliated with broker k (i.e., X or Y) who faces a state of the book s, and is indifferent between submitting a MO or LO. Similar notation applies for a buyer. The cutoff private valuation $b_k^S(s)$ depends also on the trading protocol. Thus a state s creates different cut-off values between PT and PBT.[13] We denote the state of the book s by $(\mathbf{q}_1^{\mathbf{i}}, \mathbf{q}_2^{\mathbf{j}})$, where $\mathbf{q}_1^{\mathbf{i}}$ is a vector that represents the orders standing at the bid and the superscript

[11] We assume that the arriving investor has complete information about s, i.e., whether there are orders in the book, for how long they are in the book, and by which broker they have been submitted. In Section 8, we relax some of these assumptions.

[12] This cost can be associated with the effort that brokers make in order to monitor the market (see, e.g., Foucault et al., 2013) and we make the assumption that in expectation is driven by a zero profit condition. In Subsection 5.2 we explicitly model a free-entry equilibrium where dealer-specialists have a cost c per time period to make the market. Dealer-specialists then do not enjoy equilibrium profits and hence are not relevant for welfare.

[13] For example if s is the empty state, then $b_k^S(empty)$ in PT is different from $b_k^S(empty)$ in PBT.

$\mathbf{i}$ denotes the periods for which orders have been in the book as well as the broker affiliation. Thus for example, $\mathbf{i} = (1, k)$ or $(2, k)$, $k \in \{x, y\}$ implies that the order is in the book already for one or two periods, respectively, and was submitted by a trader affiliated to broker k. For notational simplicity we write $\mathbf{i} = k$ instead of $\mathbf{i} = (1, k)$, $k \in \{x, y\}$, for limit orders standing for one period. Similar interpretation is used for $\mathbf{q}_2^{\mathbf{j}}$. Following the literature, we denote standing orders at the bid with a positive sign and and at the ask with a negative sign. To illustrate, we provide a few examples: $(0, 0)$ denotes an empty book on both sides; $(0, -1^y)$, a book with no limit orders at the bid and a limit order standing at the ask for one period submitted by a seller affiliated to broker Y; $([1^{2,x}, 1^x], 0)$, a book with two limit orders standing at the bid for two and one periods respectively both submitted through broker X, and no limit orders at the ask. Figure 1 shows the lifespan of an investor and the timing of her actions.

*** **Please Insert Figure 1 about here*****

Note that under the two period cancellation rule, not all states are feasible. For example $[(1, 1), -1]$ is not a feasible state regardless the broker affiliation and the number of periods standing, because it would require for orders to stay in the book for, at least, three periods. The proposition below and its proof follow from the discussion above.[14]

Proposition 1 *A seller's cut-off values b_k^S, $k \in \{x, y\}$ depend on her broker affiliation with PBT, the ask and bid prices and the state s of the book at her own side. Given the two-period life span of LOs, they do not depend on the state of the book at the opposite side.*

Proposition 1 is a consequence of our model's assumptions, and in particular the two period lifespan of a limit order and the existence of a dealer-specialist.[15] The importance of Proposition 1 is that it allows us to limit the number of relevant states for each arriving investor. In particular a seller, will form her decision endogenously but independently of the state of the book at the bid.

[14]The symmetries of our model allows to focus on sellers. All results can be reproduced for buyers.

[15]The model can be extended to more periods, but it does not add to intuition and at the same time makes notation and calculations more cumbersome. The result that the cut-off values do not depend on the state of the book at the opposite side is particular to the two-period lifespan.

The cut-off values reflecting indifference between submitting a MO or a LO are a function of the heterogeneity in investors' valuations ($2V$), the tick and the probability of execution. In particular we have the following proposition.

Proposition 2 *Let b_k^S, $k \in \{x, y\}$ be the cut-off value that makes the seller indifferent between a MO and a LO. Let P denote the probability of execution of the LO, Δ the tick size, and $2V$ denote the heterogeneity in investors' valuations of the asset. Then*

$$b_k^S = 1 - \frac{Z_0 - \Delta}{2V} f(P),$$

where f is an increasing function with respect to the price and given by

$$f(P) = \frac{1+P}{1-P}.$$

We note that in the above proposition the probability of execution depends on the state of the book i.e. $P = P(s)$. As the probability of execution P increases, the higher the incentives of the trader to submit a LO as opposed to a MO. That is reflected in the decrement of b_k^S caused by f. We also notice that the strategy of the agent is affected by the *relative tick*, i.e., the tick relative to the heterogeneity in traders' valuation, which equals in our setting $2V$. As the relative tick increases, the benefits from submitting a MO decrease and thus b_k^S decreases.

We say that a state of the book s' is irrelevant to s, if $b_k^S(s) = b_k^S(s')$, $k \in \{x, y\}$. Otherwise the states are relevant for the trader. For a given state s of the book, we denote by $-s$ its symmetrical. For example if $s = (1^{2,x}, 0)$ then $-s = (0, -1^{2,x})$.

Proposition 3 *Let $b_k^S(s)$, $(b_k^B(s))$, denote the cut-off value that makes a seller (buyer) indifferent between submitting a LO or trading via a MO, when the state of the book is s. Then for all $k \in \{x, y\}$ the following hold:*

(i) $b_k^B(s) = 2 - b_k^S(-s)$.

(ii) $P(b \geq b_k^B(s)) = P(b \leq b_k^S(-s))$.

2.2 The arriving trader's decision problem

This subsection discusses the decision problem for an arriving seller for the two trading protocols. We consider first, the case of PT. Using Proposition 1, we deduce that only two groups of states are relevant: one group comprises states in which she faces no competition in the LOB and constitutes of the empty book on her side and the irrelevant states to that, and the second group contains the states that place her second-in-line on her side as well as its irrelevant states to that, i.e., a LO standing that does not expire immediately.[16] Any limit order submission which creates an irrelevant, for the seller, to an empty book state, will result in forming the same decision as if the book was completely empty on her side. Similar reasoning holds when the trader faces competition on the book.

Under PBT, the relevant states are augmented by one as a seller now distinguishes between having competition by a broker of same or opposite affiliation. In the following we use interchangeably the terms 'no competition' and being 'first-in-line', for a trader that faces an empty book state (or irrelevant to that) upon arrival to the market. Respectively, we say that a trader faces 'tough' ('soft') competition if the book contains queue, formed by a trader with the same inclination and employing the same (different) broker. For competition under PT, we use the term 'intermediate' competition to describe the states in which the seller finds non-expiring queue at the book formed by traders with same inclination.[17]

We refer to Internet Appendix A for details about the system of equations (System 5) that needs to be solved for PT and PBT.

At this point, a discussion on the execution probability of a limit order is in order. Notice that, *ceteris paribus*, a seller arriving on the market who submits a limit order, has higher probability of execution when facing an empty book on her side rather than joining the queue. Respectively, for any particular state that leads to joining the line, if the trader is subject to preferential execution then her execution probability increases.

[16]Table D1 in Internet Appendix D depicts them in detail. We note that a trader in PT is not interested in distinguishing between *same* and *opposite* broker affiliation.

[17]If it is clear by the content, for PT and PBT we may write that a *trader is second-in-line*, for a trader that faces competition upon arrival to the market.

Given this remark, we are able to formulate the following proposition, which identifies the relation between the cut-off values.

Proposition 4 *Assume an x-seller arrives at the market, then the following holds:*

i) $b_x^S(0,0) < b_x^S(0,-1^y) < b_x^S(0,-1^x)$*, under PBT.*

ii) $b_x^S(0,0) < b_x^S(0,-1)$*, under PT.*

iii) $b_x^S(0,0)$ *under PT is less than* $b_x^S(0,0)$ *under PBT.*

iv) $b_x^S(0,-1)$ *under PT is larger than the average of* $b_x^S(0,-1^y)$ *and* $b_x^S(0,-1^x)$ *under PBT.*

Similar relations for the cut-off values of a y-seller hold. Proposition 4 provides an insight on the behavior of traders when facing a queue. According to our intuition, we observe that the longer the queue that a trader faces, the more aggressive she becomes and the more likely she is to submit a MO. This is in accordance with both theoretical and empirical findings (see, e.g., Parlour, 1998; Ranaldo, 2004). In addition Proposition 4 reflects the different value standing LOs may have depending on the position on the queue and the trading protocol (see, e.g., Dahlmström et al., (2017) for the value of a LOs position in the queue).

In equilibrium, our model endogenously determines the behavior of traders. We are able to identify systematic patterns in the actions of arriving agents even though they arrive independently and have random draws from a uniform distribution related to their personal valuations. Since these actions are determined by the length and the structure of the queue as well as the expected behavior of future arriving agents, the systematic patterns differ according to the secondary priority rule that is in place. This is a novel result and in accordance to Parlour (1998) or Degryse et al., (2009), and creates empirical and testable implications also in our multimarkets interpretation.

We solve System 5, for PBT and PT, and obtain the cut-off values for the arriving sellers which define endogenously their actions. These are function of the minimum tick and thus the ask and bid prices. Let $\phi_t(\beta, s)$ denote the strategy of the arriving trader, given her willingness to trade β and the state of the book s that she faces. The following proposition formulates the empirical predictions of our model which are related to the cut-off values and hold for all magnitudes of the relative tick.

Proposition 5 *PBT predicts different systematic behavior patterns in comparison to PT. In particular the following hold:*

i) There is a higher likelihood at t+1 to observe a limit order at the ask with PBT rather than PT, if the action at t was a limit order at the ask, i.e.,

$$P[\phi_{t+1}(\beta, s) = -1| -1, PBT, t] > P[\phi_{t+1}(\beta, s) = -1| -1, PT, t].$$

ii) There is a higher likelihood at t+1 to observe a market order (transaction on the LOB or dealer-specialist) at the ask in PBT rather than PT, if the action at t was a transaction i.e.,

$$P[\phi_{t+1}(\beta, s) = Tr^A|\ Tr^A, PBT, t] > P[\phi_{t+1}(\beta, s) = Tr^A|\ Tr^A, PT, t].$$

iii) Assume at t there was a limit order submission at the ask. Then it is less likely at t+1 to observe a limit order from the same broker in PBT rather than PT, i.e.,

$$P[\phi_{t+1}(\beta, s) = -1^x| -1^x, PBT, t] < P[\phi_{t+1}(\beta, s) = -1^x| -1^x, PT, t].$$

iv) Assume at t there was a limit order submission at the ask. Then it is more likely at t+1 to observe a limit order from different broker in PBT rather than PT, i.e.,

$$P[\phi_{t+1}(\beta, s) = -1^x| -1^y, PBT, t] > P[\phi_{t+1}(\beta, s) = -1^x| -1^y, PT, t].$$

v) In PT, is more likely to have a reversal to a market order (on the LOB or against dealer-specialist) on the same side after the submission of a limit order, rather than in PBT, i.e.,

$$P[\phi_{t+1}(\beta, s) = Tr^A| -1, PBT, t] < P[\phi_{t+1}(\beta, s) = Tr^A| -1, PT, t].$$

Similar formulations hold for the bid side. Further empirical predictions of the model are presented in Section 6. From (*i*) of Proposition 5, we expect that under PBT longer queues of LOs (i.e., *depth of two*) are formed in the LOB. Also with PT it is more likely that there will be depth in the book. A more extensive investigation on the depth of the book is presented in Section 3.3.

2.3 Model Equilibrium

In this section, without loss of generality, we normalize the fundamental value V to one and we solve System 5 (please, see, Interenet Appendix A, for the complete formulation), for all possible relative tick sizes.[18] In our analysis we include both PBT and PT. As both brokers have identical market shares, the behavior of x and y traders are symmetrical and hence, we report results for a seller affiliated to broker X. Figure 2 illustrates the cut-off values for a seller for all relative tick sizes.[19]

*** **Please Insert Figure 2 about here*****

Figure 2 reveals that an arriving trader when facing 'an empty state' is more aggressive under PBT than PT. This increment reflects the anticipation effect that being first-in-line (and thus not facing competition from an ex-ante point of view) is subject to losing her line priority by a following opposite-broker trader who may perform queue jumping. PBT has not a uniform effect on traders that face competition. It incentivizes traders to join the line after an opposite-broker trader, since they may queue jump later on (i.e., enjoy preferential execution). However, 'second-in-line' traders behind a same-broker trader, have less motives to join the queue since they cannot perform queue jump, but still in the following periods, may lose their line. Proposition 4 has identified this particularity; a trader's decision to join the line is affected by the structure of the queue. In Figure 2, we report how cut-off values for a seller vary across relative ticks. Panel C identifies the difference in aggressiveness of a trader according to the competition that she faces. In PBT a trader under competition, trades always more with LOs. The opposite holds for traders who are on an empty or irrelevant to empty state. Using the cut-off values, we are able to calculate the transition matrix $\mathcal{M}$, and obtain the steady

[18]We solve numerical our model for multiple values of fundamental prices and granular tick sizes, essentially providing *proof by exhaustion or brute force.* To put things into perspective, let us assume a \$500 value stock traded under a 1 cent tick. This is equivalent to a \$1 stock traded under $2 * 10^{-5}$ tick increments. In our methodology we obtain the critical values for traders valuations with an error less than 10^{-16}, and we implement for the normalized to one stock, tick increments of 10^{-12}, much finer than any tick increment observed in today's financial markets. So our results, even though are not derived for a continuum of tick increments, still constitute a proof because they exhaust all potential increments that a \$1 dollar asset can have.

[19]Assume $V = 1$, so the tick ranges in the interval $[0, 2]$. Equivalent the half tick ranges between $[0, 1]$.

state probabilities of our system.

*** **Please Insert Figure 3 about here*****

Figure 3 shows the consolidated steady states for all relative tick sizes under PT and PBT. In Panels A and B we decompose the states that lead to 'no-competition', to the 'empty state' and to the states 'irrelevant to empty'. However, as the tick increases, queues start to form. These results in the increase of the likelihood in observing competition states. Panel D shows the difference between PBT and PT in the probability of observing competition for the arriving investor. We notice that the result is not uniform. For small ticks, it is more likely to observe competition in PT. The opposite holds for wide ticks. This compositional change will be one of the main topics of our next section.

3 Priority Rules: Market Quality and Welfare

We now study the impact of PT and PBT on market quality and investor welfare. We capture market quality through trading rates, depth of the limit order book, and fill rates of limit orders. Welfare is quantified as investors' gains from trading. Dealer-specialists make zero profits and therefore do not directly contribute to welfare.

3.1 Trading Rates

3.1.1 Unconditional and Conditional Trading Rates

We define the unconditional trading rate, hereafter trading rate, as the likelihood that any agent (investor, dealer-specialist) participates in a transaction. This is equivalent to defining the trading rate as the likelihood that the arriving investor, i.e., a seller, submits a MO that either executes against standing LOs or against the dealer-specialist.[20] The computation of trading rate, not only allows us to estimate the transacted volume, but also is essential for computing the generated welfare from trading (for details, see

[20]This trading decomposition, regarding the counterparties involved in a transaction, depends on both the tick size and the trading protocol, and has significant implications for the generated welfare, since it leads us to deviate from the notion that higher trading rates are always associated with higher generated social welfare.

Section 3.2). For the calculations we identify the probability that a trader submits a MO in each state of the book based on the stationary distribution. With a slight abuse of notation we write $P(MO)$ for the more correct $P(MO|\ seller)$.[21] Let S denote the set of all possible states s of the book indexed over a set I. Then the trading rate is defined as follows:

$$TR = P(MO) = \sum_{i \in I} P(MO|\text{ state } s_i)P(s_i).$$

Proposition 5 showed that investors' behavior depends on the 'competition' that they face on the LOB. To obtain a better understanding of the trading rate, we decompose the states of the book that place the seller to an 'empty state or first-in-line' and to those under which she faces 'competition or second-in-line' upon arriving. Let $B_F, (B_S)$ be subsets of indexes of I for which the arriving trader is first-in-line (second-in-line) and thus faces no competition (competition) when she arrives to the market, and let $TR_F, (TR_S)$ denote the corresponding trading rate, then

$$TR_F = \sum_{i \in B_F} P(MO|\text{ state } s_i)P(s_i), \text{ and } TR_S = \sum_{i \in B_S} P(MO|\text{ state } s_i)P(s_i),$$

and it holds that $TR = TR_F + TR_S$.

Next, we define the conditional trading rate for the first-in-line and second-in-line as the probability that a trader will submit a market order, given her position on the line, i.e.,

$$P(MO|k \text{ in line}) = P(k \text{ in line})^{-1} \sum_{i \in B_j} P(MO|\text{ state } s_i)P(s_i),$$

where $(k, j) \in \{(first, F), (second, S)\}$.

In PT, both the unconditional and conditional trading rates, for a second-in-line trader are irrelevant to the structure of the queue. However, in PBT, since the likelihood of execution depends on whether the trader faces 'soft' or 'hard' competition we have

[21] We calculate the trading rate of an arriving seller and not of an x- seller, unless explicitly stated. Since sellers arrive through brokers with equal market shares, we obtain that $P(MO|\ seller) = P(MO)$.

that the investor's broker affiliation is also essential to her decision. Therefore, under PBT, the unconditional and conditional trading rates depend also on the particular formation of the queue.

3.1.2 Trading Rates and Relative Tick Size

Our model is able to study heterogeneity stemming from variation in the relative tick size (i.e., $\Delta/2V$).[22] In the two extremes in which Δ equals 0 or $2V$ (i.e., the maximum heterogeneity in investors' willingness to trade), PT and PBT coincide. This happens since in the first case traders trade only via MOs against a dealer-specialist and hence there is no queue accumulation in order to observe the structural differences of PBT.[23] In the second case when the tick equals the maximum heterogeneity in investors' willingness to trade, the market collapses as all investors abstain from trading.

More generally, when the relative tick size is small, we expect to observe higher trading rates with PBT rather than PT. With small ticks, the 'empty state' occurs quite frequently. With PBT and faced with the empty state, investors anticipate that when submitting a LO, they face the possibility of losing their line when later arriving investors from another broker joing the queue. This *anticipation effect* induces them to act more aggressively by submitting more often MOs (either against standing LOs or against a dealer-specialist). As a result we have a higher trading rate. In contrast, when the relative tick becomes wide and therefore more investors participate in the market, queues start to form, such that traders when arriving to the market, frequently face the 'competition' state. Under PBT these traders have more incentives to join the queue (i.e., *queue-joining effect*) in particular when faced with 'soft competition'.

In sum, when the relative tick increases, we expect the positive difference in trading rate between PBT and PT to decrease and after a critical point to reverse. We define $\tilde{\Delta}$ to be the critical tick size for which trading rates under PT and PBT are identical,

[22] We denote here the tick size relative to the heterogeneity in investors valuation ($2V$). The tick size relative to the common value of the stock would be $\Delta/(Z_0 + V)$

[23] In that case, we would also need that dealer-specialists face no costs in making the market.

with $\tilde{\Delta} \approx 0.89868$. We refer to any tick Δ smaller (larger) than $\tilde{\Delta}$ as tight (wide).[24]

The following Proposition summarizes the findings on trading rates as a function of the tick size. Define Δ as the tick size, and $\tilde{\Delta} \approx 0.89868$.

Proposition 6 *If $\Delta < \tilde{\Delta}$, the trading rate is higher with PBT in comparison to PT. If the tick is larger than $\tilde{\Delta}$, the opposite holds.*

Figure 4, Panels A and B, illustrate the trading rates for PT and PBT as a function of the size of the half-tick (which varies between 0 and 1), respectively. These panels also show the trading rate by traders being first-in-line (TR from FiL) and second-in-line (TR from SiL). Panel C illustrates the difference in trading rates between PBT and PT. Notice that the trading rate equals 50% when the half-tick is zero whereas it becomes 0% when the half-tick equals 1. For both trading protocols, the total trading rate and the trading rate from FiL investors drops as the half-tick widens. The trading rate from second-in-line investors shows an inverted U-shape as a function of the half-tick for both trading protocols. We further observe that the trading rate becomes higher in PT than in PBT when ticks are tight. In contrast, the trading rate is higher with PBT than PT when ticks are wide.

Panel D of Figure 4 shows the difference in *conditional* trading rates between PBT and PT. When there is 'no competition', conditional trading rates are higher with PBT reflecting the anticipation effect. When there is 'soft' competition, traders are more willing to join the queue with PBT than with PT whereas the opposite holds for 'hard' competition. Overall, however, the impact of 'soft' competition dominates the one on 'hard' competition, leading to the queue-joining effect.

*** **Please Insert Figure 4 about here*****

[24]For an arbitrary support of the distribution β, the critical tick size is approximately equal to $0.44934 \cdot V(b_{max} - b_{min})$, where b_{max} is the maximum and b_{min} the minimum of the willingness to trade distribution.

3.2 Investor Welfare

We compare investor welfare for our two trading protocols PBT and PT. Our measure of ex-ante investor welfare is based on the behavior of rational traders and therefore is on average equal to the ex-post. In our model, dealer-specialist realize zero profits and therefore do not directly contribute to welfare. We first introduce our definition of investor welfare for trades in the LOB and against dealer-specialists. Our measure W captures investors' gains from trading as in Colliard and Foucault (2012), or Parlour and Seppi (2003). We denote by $\pi(\phi_t(\beta, s))$, the investors' gains from trading at t, following the strategy $\phi_t(\beta, s)$. We note that the strategy depends on her private valuation of the asset, measured by β, and on the state of the book s. Hence,

$$W \equiv E(\pi(\phi_t(\beta, s))).$$

The expectation is taken with respect the product space created by the willingness to trade β and all the steady states of the book s. We calculate the welfare generated in each of the intervals created endogenously by the cut-off values, as shown in Figure 2, and let $\pi_i(\phi_t(\beta, s))$ denote the welfare generated over the A_i interval. In particular, in PBT, there are three relevant states, each creating a cut-off value that differentiates the decision of the arriving trader between a MO and a LO (or refrain from trading), given the state of the book.[25] The cut-off values decompose the interval of potential valuations $[0, 2]$, to four sub intervals, which allow us to evaluate investor welfare as follows:

$$W = \sum_{i=1}^{4} E(\pi_i(\phi_t(\beta, s))).$$

We denote by I, the set of indexes that enumerate all states of the book S. Then we further analyse welfare over the interval A_i as:

$$W_i \equiv E(\pi_i(\phi_t(\beta, s))) = \int_{A_i} \sum_{j \in I} (w_j \cdot P(s_j)) dF_\beta,$$

[25]For instance, if the book is empty, valuations close to zero lead to MOs and hence $A_i = [0, \widetilde{b}]$, where $\widetilde{b}$ is determined endogenously.

where w_j depends on the interval on which the welfare is evaluated as well as the state of the book and represents the generated gains. If, for example, (b_j, s_j) is such that it would lead the arriving seller to submit a MO then w_j is equal to $(B - Z_0 - b_j V)$ while if she opts for a LO then $P(LO_j)(A - Z_0 - b_j V)$, where the subscript j, denotes that the execution probability of the LO depends on the state of the book s_j.

Figure 5, Panels A and B, show welfare as a function of the size of the half-tick for PT and PBT, respectively. Not surprisingly, welfare is highest when Δ equals zero. When the half-tick equals one, welfare becomes equal to zero as no trading gains can be realized for any possible investor match. For these two values, PT and PBT coincide in terms of generated welfare. We notice that when the half-tick increases, welfare continuously decreases for both trading protocols. Panel C of Figure 5 shows the difference in welfare between PBT and PT as a function of the half-tick. When the half-tick is smaller than $\tilde{\Delta}/2$, PBT generates lower welfare than PT, whereas the opposite holds when the half-tick is larger than $\tilde{\Delta}/2$.

Panels D to F of Figure 5 aim to explain where the difference in welfare between the two trading protocols comes from. We first argue that trader composition differs considerably between the two protocols. Panel D shows that first-in-line traders trade more often under PBT than under PT whereas the opposite holds for second-in-line traders. In Panel F, we measure the difference in the counterparties involved in a transaction between PBT and PT. There are more (less) transactions having as a counterparty investors (a dealer-specialist) in PT (PBT) when the half-tick $< \tilde{\Delta}/2$ ($> \tilde{\Delta}/2$). This shows, as expected, that the main force for welfare is transactions between traders. In order to illustrate the compositional change between tight and wide ticks, Panel F, shows the difference in trading rates for different tick sizes, by competition and by the different categories of counterparties involved. We observe that trades by first-in-line traders are higher in PT in tight spreads, but this rapidly changes in favor of PBT.

*** **Please Insert Figure 5 about here*****

Furthermore, trader composition translates into generated welfare. A sell market order that executes against another final investor generates $(b_{buy} - b_{sell})V$ in terms of welfare as both counterparties enjoy trading gains. In contrast, a similar sell market order that executes against the dealer-specialist generates only welfare for the investor who submits the market order, i.e., $B - Z_0 - b_{sell}V$. This becomes more important the lower B and thus the larger the half-tick.

Overall we observe a compositional change between PT and PBT related to welfare. This is summarized in the following proposition.

Proposition 7 *If $\Delta < \tilde{\Delta}$, then welfare is higher under PT relative to PBT. If $\Delta > \tilde{\Delta}$, welfare is higher in PBT than in PT.*

3.3 Fill Rate and the Depth of the LOB

Proposition 5 identifies systematic trading patterns. This has important implications to fill rates and the depth of the LOB. The fill rate of a LO is defined as the execution probability of a submitted LO, and has been studied both theoretically (Colliard and Foucault, 2012) and empirically (Malinova and Park, 2015). Depth is a major characteristic of market quality, and refers to the ability of the LOB to sustain large market orders without impacting the price (Kyle, 1985). We note that the higher the fill rate, the higher the incentives for arriving investors to submit a LO and therefore actively contribute to the making of liquidity. Thus, higher fill rates imply a more liquid market. We define the fill rate of a LO as follows:

$$FR = \frac{\sum_{i \in I} P(s_i) P_S(\text{submission of } LO_i) P_S(\text{execution of } LO_i)}{\sum_{i \in I} P(s_i) P_S(\text{submission of } LO_i)},$$

where $P_S(\text{submission of } LO_i)$, ($P_S(\text{execution of } LO_i)$) denote the likelihood of submission (execution) of a limit order when the state of the book is s_i.

Panels A to C of Figure 6 show our findings on fill rates. Panels A and B show the fill rates as a function of the half-tick for PT and PBT, respectively. We notice that fill rates of LOs drop in the half-tick for both trading protocols. Panel C shows the difference in fill rate between PBT and PT as a function of the half-tick. We observe that for $\Delta < \tilde{\Delta}$, the fill rate for PBT is lower than PT whereas the opposite holds for

$\Delta > \tilde{\Delta}$. The fill rates are positively correlated with the trading rates among investors as these lead to the execution of LOs.

We now compare the average depth ($\overline{D}$) supplied by traders under the different trading protocols as well its variation, between no depth, depth of size 1 or size 2. We define the average depth as a weighted average, using as weights the Markov stationary distribution.

$$\overline{D} = \sum_{i=0}^{2} P(\text{ States Creating a queue of } i) \cdot i.$$

The effects of PBT relative to PT on depth vary. PBT incentivizes traders that face soft competition to join the queue, resulting in 'depth of 2'. However, the anticipation effect may lead to lowering the 'depth of 1'.

*** **Please Insert Figure 6 about here*****

In Figure 6, Panel D we report the average depth in PBT as well as the variation in depth on one side of the book as a function of the half-tick. We notice that the average depth increases in the half-tick as more traders participate by submitting LOs. While 'depth of 2' increases in the half-tick, 'depth of 1' exhibits an inverse U-shape. The latter stems from the queue-joining effect as it becomes more and more attractive to join the queue when the tick is wide. Panel E displays the difference in depth between PBT and PT. We observe that under any tick size the average depth under PBT is lower than under PT. The anticipation effect leads to a lower 'depth of 1' and the queue-joining effect to a greater 'depth of 2' for all sizes of the half-tick. With PBT, there are transitions where 'depth of 2' fades towards no depth when queue-jumping and order expiration takes place.[26] Finally comparing Figures 5 and 6, we notice that in large ticks, welfare increases and depth decreases. This seemingly inverse relation should not come as a surprise, since the higher welfare observed in large ticks is mainly driven by investor-investor trades, which by definition decreases depth.

[26]Under PT, the likelihood of observing an empty book after 'depth of 2' the following period is 0.

4 Endogenous Adoption of Priority Rules

This section investigates which trading protocol each of the two brokers wants to adopt when given the choice. We assume that the adoption of PBT or PT is (i) a one time decision before the game starts, and (ii) each broker maximizes the welfare of its clients anticipating their order submission strategies. While brokers could possibly charge a fee when adopting one priority rule over another, we start from the presumption that Bertrand competition would drive these fees to zero when investors ex-ante could decide on which broker to affiliate with.[27]

To study the endogenous adoption of priority rules, we also have to investigate the case where one broker offers PT whereas the other offers PBT. A unilateral adoption of PBT by a single broker, offers the benefit of obtaining preferential execution on behalf of his clients, while not being subject to losing their position in the queue by a future arriving trader affiliated with the opposite broker. In order to model it, we modify our main system of indifference equations, presented in Internet Appendix A.[28]

Proposition 8 *Assume that brokers can unilaterally decide whether to offer PBT or PT in order to maximize their investors' welfare before the trading game starts. Then under any tick size, the 2-by-2, non-cooperative, symmetric game has a unique Nash equilibrium, in which both brokers decide to offer PBT.*

Figure 7, forms the proof of Proposition 8 and plots the difference in welfare between PBT and PT, obtained by the clients of broker X given that broker Y adopts PT (Panel A) and PBT (Panel B) for all sizes of the half-tick. For a 2-by-2 symmetrical game, since the differences are positive for all tick sizes, we obtain that the unique Nash-equilibrium is found when both brokers adopt PBT. We notice that for ticks below $\tilde{\Delta}$, brokers are in a prisonner's dilemma: they would be better off if they could commit to implement PT. For ticks larger than $\tilde{\Delta}$, the adoption of PBT also jointly maximizes all investors' welfare.

[27]Even when brokers charge fees, similar results would apply to the extend these fees are positively correlated with their clients welfare.

[28]The main alteration is that in these indifference equations, one broker offers preferencing and the other not. That affects the execution probabilities of LOs, and therefore the Markov steady state equilibrium.

*** **Please Insert Figure 7 about here*****

Our analysis thus shows that even though PBT results as a Nash-equilibrium, it does not coincide with the social optimum when the tick is small. The private and socially preferred outcome differ as an individual broker does not internalize the negative impacts its preferred priority rule has on the other broker's investors.

5 Priority Rules: market share and number of dealer-specialists

5.1 Priority rules: market share and trade composition

Priority rules affect the market shares of customer-customer trades (i.e., trades on the LOB involving investors on both sides of the trade) and customer-dealer/specialist trades (i.e., trades where a dealer-specialist is involved at one side of the trade). This follows from our discussion on trade composition (Panel E of Figure 5).

Figure 8 shows the market shares for customer-customer trades as a function of the half-tick for both trading protocols. The complement is the market share of customer-dealer/specialist trades. We find that the market share of trades on the LOB (i.e., customer-customer trades) increases in the size of the half-tick for both trading protocols. Dealer-specialist involvement thus decreases in the size of the half-tick. Hence, a tight tick is associated with an important intermediation function by dealer-specialists as few investors prefer to submit a LO over a MO in the presence of dealer-specialists. With tight ticks, the rate of MOs is high, such that there are few LOs in the book, implying a great involvement of dealer-specialists. Further, the market share of customer-customer trades is greater (smaller) with PBT than PT when the tick is greater (smaller) than $\tilde{\Delta}$.

*** **Please Insert Figure 8 about here*****

Priority rules also influence the *composition* of customer-customer trades, i.e., the share of trades where the counterparties use the 'same-broker' versus 'different-brokers'.

With PT, the structure of our model implies that 'same-broker' and 'different-broker' trades are equally important. With PBT, the 'same-broker' trades outweigh 'different-broker' trades stemming from queue-jumping. Figure 9 shows how the share of 'same-broker' trades as a fraction of all customer-customer trades evolves as function of the half-tick. We observe that the share of 'same-broker' trades increases in the tick size implying that more queue-jumping occurs as the tick becomes larger. Intuitively, with larger ticks, more queues are formed leading to bigger opportunities to queue jump.

*** **Please Insert Figure 9 about here*****

5.2 Priority rules and number of dealer-specialists

Finally, we show how we can endogenize the number of dealer-specialists and how this number depends on priority rules in the LOB. Assume that a dealer-specialist has a cost c per time period to organize and monitor the market. We assume c to be independent of priority rules. Dealer-specialists are not facing inventory costs and are risk-neutral. When there are several dealer-specialists in the market, we assume a trader randomly chooses one of them. A dealer-specialist would stay and make the market as long as its expected profit per period exceeds its cost, i.e.,

$$[2\ (1 - \frac{B - Z_0}{V})\ TR^i_{Specialist}]^{-N_i} \geq c, \tag{1}$$

where $TR_{Specialist}$ corresponds to the trading rate against dealer-specialists, N_i is the number of dealer-specialists and $i \in \{PBT, PT\}$. We assume competition among dealer-specialists implying that the above inequality becomes binding for both trading protocols. This setting also implies that dealer-specialist do not make profits and do not directly contribute to welfare. The implication then is that we have more dealer-specialists when the tick is tight as the trading rate against dealer-specialists is high. Further, when comparing priority rules, the implication of Panel E of Figure 5 is that we have more (less) dealer specialists with PBT than PT when the tick is smaller (greater) than $\tilde{\Delta}$.[29]

6 Empirical and Regulatory Implications

In Proposition 5, we identified empirical implications of our model, and the effects of priority rules on trading rates and depth have been discussed in Sections 3.1.1 and 3.3. Here, we formulate testable implications which can be of use to empirical researchers. We build on the empirical predictions provided in Parlour (1998) which were stated under PT environment.[30] We first identify implications that are specific to PBT. Next we focus on differences in systematic order flow patterns across the two priority rules.

In PBT, a seller under soft competition has more incentives to join the line rather than when she faces tough competition. This is due to the probability of queue jumping. An implication is that the likelihood that two consecutive LOs at the ask coming through the same broker, would be lower as opposed if they were submitted by traders affiliated to different brokers. This implication is specific to PBT and is reflected in the relation that the cut-off values have, provided in Proposition 4. Thus it depicts the systematic trading patterns *within* PBT.

Testable Implication 1 *In equilibrium under PBT and at one side of the book, limit orders are more likely to be followed by limit orders coming from traders affiliated to different brokers in comparison to being affiliated to same brokers.*

From the above implication follows that under PBT, reversion to a MO after a LO is more likely to occur by traders having the same broker affiliation.

Next, we turn our focus on the systematic patterns created *between* the two trading protocols. Traders facing competition are more reluctant to join the queue as opposed to those who do not. This is true for both trading protocols, but with different intensities. In PBT, competition incentivizes traders under opposite forces. In tough competition,

[29] When c is too large and Equation 1 is not satisfied for any N, we do not have the presence of a dealer-specialist. This case is examined in Section 8.

[30] In Parlour (1998), the trading patterns analyzed also refer to the opposite side of the LOB. In our model this is not the case due to the two period cancellation rule. Adding one more period to the lifespan of limit orders would allow us to formulate similar claims.

a trader becomes more aggressive in order to compensate for the possibility of losing her position on the queue in the future. Under soft competition, she is more inclined to submit a LO and exploit the possibility of a preferential execution. However, relative to PT, the dominant force is the latter one.

Testable Implication 2 *In equilibrium under PBT and at one side of the book, limit orders are more likely to be followed by limit orders than under PT.*

Next, we focus on the relation between transactions observed and limit order submissions. From our model we obtain that if a trader transacts at one side taking liquidity and reducing the queue, then the arriving trader will exploit this on her favor. Hence limit orders on the ask, are positively correlated to market orders at the ask, transacted at the previous period. The next testable implication depicts this relation.

Testable Implication 3 *In equilibrium under PT and at one side of the book, market orders are more likely to be followed by limit orders than under PBT.*

The next implication is of high importance to regulators, and identifies the optimal response of brokers regarding whether or not to introduce PBT.

Testable Implication 4 *In equilibrium, under the assumption that brokers maximize their investors' welfare, if given the option, then they will adopt PBT over PT.*

This testable implication is in line with what was observed when Euronext in 2007 following new regulation offered the possibility to its brokers to opt for PBT (through its internal matching service) or PT. Almost immediately, a vast majority of brokers signed up for the internal matching service.

The effects of priority rules to depth of the market were investigated in Section 3.3. In PBT, queue jumping with order cancellation and traders that face no competition affect the overall depth of the book and are the main driving force of the next implication.

Testable Implication 5 *In PT, the average depth of the book provide by investors is greater in comparison to PBT. The likelihood of observing a queue of 'depth of 2' is* *larger in PBT than in PT. However, this depth may reverse to an empty book more rapidly in PBT than PT.*

Our analysis on priority rules also has regulatory implications. We show that PT is socially preferred when the tick is tight but PBT is socially preferred under wide ticks. The endogenous market outcome stemming from brokers' priority rules decisions differs from the socially preferred ones when the tick is small.

Testable Implication 6 *When regulators have the choice between PBT and PT, the socially preferred outcome is requiring PT when the tick is small. For large ticks, regulators prefer PBT.*

7 Centralized versus Fragmented Markets

In the introduction we put forward an alternative 'two platform interpretation' of our model. In our setting, imposing PT across the two platforms coincides with a *centralized* market as all orders across the two platforms follow time priority. PBT implies *fragmented* markets where each platform has a group of traders that exhibit price-platform-time priority.

We now add the main testable implications that are unique to such a *centralized* versus *fragmented* markets interpretation. These implications are related to order flow for both limit and market orders. The first set of testable implications differ *between* market design choices. In fragmented markets, since traders have the opportunity to 'jump the queue' as some investors have a preference for one platform over another, we expect to have a larger persistence in limit orders than in centralized markets. The opposite holds for market orders.

Testable Implication 7 *In equilibrium, in fragmented markets, limit orders are more likely to be followed by limit orders, than in centralized markets.*

Testable Implication 8 *In equilibrium, in centralized markets, market orders are more likely to be followed by limit orders, than in fragmented markets.*

Our model also provides a testable implication for order flow *across* the different platforms when markets are fragmented (i.e., investors have price-platform-time priority). Limit orders on one platform are more likely to be followed by limit orders on the other platform than on the same platform.

Testable Implication 9 *In equilibrium, in fragmented markets, limit orders are more likely to be followed by limit orders in different platforms in comparison to the same platform.*

This testable implication shows that the composition of order flow on one platform is not only influenced by the state of the LOB on that platform but also by the state of the LOB on the other platform. This shows that it is important to analyze the market as a whole rather than focusing on specific platforms in isolation.

Testable Implication 10 *Even with price priority, fragmentation across multiple markets endogenously occurs in the absence of time priority across markets, and brokers having as tie-breaking rule a preference for one market.*

The order protection rule (i.e., trade-through prohibition) thus leaves incentives for the creation of multiple markets as time priority is not enforced across markets. This provides theoretical support for market fragmentation across different platforms that we observe many countries (e.g., Degryse et al., 2015).

8 Extensions

8.1 Trading under opacity

Our main model assumed that traders are informed about broker affiliations of standing LOs. We relax this to study opacity: traders observe whether there are standing LOs and for how long, but are not informed about the counterparty's broker affiliation. With opacity, arriving agents only observe the depth of the LOB and when standing orders expire, but not broker affiliations of these orders. Compared to transparency, opacity about a standing LO's broker affiliation only plays a role under PBT as with PT broker affiliation is irrelevant for the arriving investor's decision. With preferencing, investors form expectations on the broker affiliations of standing LOs in the LOB. We denote by PBT-O, a LOB that operates under PBT priority rules and opacity. We keep all other assumptions as in our main analysis. As with our model presented in Section 2, we construct a system of indifference equations which depict the trade-off between a market and a limit order.[31] This system yields cut-off values where agents are indifferent between submitting a market or limit order, and a limit order or refrain from trading. The agent forms her decision based on her private valuation and on her place in the queue, i.e., she has only two relevant states. The first two parts of Proposition 9, compare the relation between PBT-O and PT, while the other parts between PBT-O and PBT.

Proposition 9 *For any tick size the following holds:*

i) A seller facing 'no competition' under PBT-O is more likely to submit a MO, rather than under PT.

ii) A seller facing 'competition' under PBT-O is less likely to submit a MO, rather than under PT.

iii) A seller facing 'no competition' is less likely to submit a MO under PBT-O rather than under PBT.

iv) A seller facing 'competition' is more likely to submit a MO under PBT-O rather than under PBT.

Our results for PBT-O are intuitive and can somehow be considered as softening the impacts of PBT. PBT-O is more beneficial for traders that face competition upon arrival compared to PT. An agent arriving to an empty book still may lose her place in the queue compared to PT. Thus she accounts for that and becomes more aggressive than

[31]The system is defined following the same methodology that we used in Section 2, and for the sake of brevity is not reported.

in PT. Opacity thus can be seen as dampening the anticipation effect and increasing the queue-joining effect as compared to the transparency case.

8.2 Broker Concentration: Number of Brokers

In our main analysis we assumed two brokers. We now study the impact of varying the number of brokers N. For tractability, we assume that each broker has a market share $1/N$. As N increases, the probability of queue jumping decreases. The reason is that it becomes less likely to attract a same-broker counterparty leading to preferential execution. As a result the effects that PBT creates, i.e., the anticipation effect and the queue joining effect, decrease in N.[32] The behavior of traders align to the behavior under PT when N goes to infinity.[33]

8.3 Priority rules and incentives for off-exchange reporting

Priority rules may impact the incentives and degree of off-exchange trading. When PT applies, a broker could circumvent it and implement PBT by reporting some trades off-exchange. In particular, a broker can induce queue jumping by matching a MO against one of its clients' LO standing deeper in the queue, and report these trades off-exchange. To the extent that this happens, we expect more off-exchange reporting when trading venues implement PT rather than PBT. The reasoning is that this force is not present with PBT as orders already queue jump on the LOB. To understand the magnitude of this force, we compute the trades stemming from queue-jumping in PBT as a fraction of the total trading rate. Figure 9, Panel B, shows that this share increases as the tick size becomes larger. Our model thus predicts more off-exchange reporting in markets with PT, a force that is more pronounced when the tick is wide. Recent evidence from the US Tick Size Pilot program is consistent with our predictions. This program was initiated in October, 2016; with a two year duration. It aimed at increasing liquidity

[32]The number of brokers and the critical tick size $\tilde{\Delta}$, are negative correlated, i.e., as N increases, the spread after which we observe higher trading rate in PT, decreases.

[33]Under a similar argument we can show that with N=2, PBT converges to PT when one of the broker's market share approaches one.

of small cap stocks by increasing the tick size. However, initial findings indicate that daily volume is unaffected but a shift towards off-exchange trading occurred. The latter is consistent with our model to the extend that brokers circumvent PT (EMA, 2017; Pearson and Ruane, 2017; Michaels, 2018).

8.4 Absence of dealer-specialists

Our main analysis assumed the presence of dealer-specialists to trade against in the absence of standing LOs. We now study what happens in the absence of such dealer-specialists. Figure 10 graphs the difference in the trading rates (Panel A) and welfare (Panel B) between PBT and PT for all half-tick sizes.

*** **Please Insert Figure 10 about here*****

Panel A of Figure 10 shows that for low half-ticks, PBT yields lower trading rates than PT whereas the opposite holds for wider half-ticks. This differs from our main model and can be explained as follows. In the absence of dealer-specialists, the possibility of trading via a MO crucially depends on the presence of a standing LO on the opposite side in the book. An investor who arrives at the market and cannot submit a MO has to decide between exiting the market or posting a LO. Every investor that would have considered a MO in the presence of dealer-specialists now has to submit a LO, implying that more LOs are present in the LOB. The act of queue-jumping with PBT then creates more often a state with an empty book leading to fewer trading possibilities compared to PT. This is particularly important when the tick is small, explaining the lower trading rate for small ticks compared to our main model. Further, compared to the main model, the anticipation and queue-joining effects thus are not at work when investors cannot choose between a LO or a MO. In states where investors have the choice, both effects still drive investors behavior.

Panel B of Figure 10 shows the difference in welfare between PBT and PT as a function of the half-tick. In the absence of dealer-specialists, trading volume and welfare co-move positively. In this setup, welfare is always generated by customer-customer trades who strongly contribute to welfare. The implication is also that in this modified setting the welfare results are in accordance with our main results obtained in Section 3.2.

8.5 Two ticks

In this subsection we discuss the case where dealer-specialists quote at one tick inferior to the best bid and ask quotes used by LO traders. This could stem from dealer-specialists having high costs of being around preventing them from making the market at the most competitive level. This setup is somehow a combination of the approach in our main model and the one without dealer-specialists as discussed in subsection 8.4.

When the tick is too wide (i.e., when $\frac{\Delta}{2} \geq \frac{V}{3}$,), the outer quotes put by dealer-specialists become unattractive to all potential traders such that the LOB dominates. This has as consequence that then investors submit LOs to the LOB, or trade against another investor in the LOB, as discussed in subsection 8.4.

When $\frac{\Delta}{2} < \frac{V}{3}$, the outer quotes of dealer-specialists start to bite, and the model becomes a combination of our main setting and the one in subsection 8.4. The anticipation effect drives the investor-dealer/specialist trading rates to be higher in PBT relative to PT. A welfare analysis of this section shows that the results obtained from our main model still hold, where PT yields higher welfare compared to PBT for low ticks and PBT is preferred for larger ticks.

8.6 Random Matching

Next to PT and PBT, other secondary tie-braking rules exist in financial markets such as random matching (RM hereafter).[34] Under RM, all standing LOs have the same probability of execution, regardless of their time of submission (see, e.g., Dugast, 2018 who employs this allocation mechanism in his theoretical model). Accordingly, RM provides strong incentives to join the queue. Since under RM, on average the likelihood

[34]Other examples include display, size, and pro-rata matching. In subsection 8.1 we already discussed opacity of broker IDs. In our setting, RM, pro-rata and size coincide for indivisible one unit orders. Pro-rata matching is particular popular within the derivatives markets (see, e.g. NYSE Liffe futures market; CME Group).

of queue jumping increases compared to PBT, intuitively we would expect that RM would have similar but more pronounced effects compared to PBT. Both the anticipation effect as well as the queue-joining effect should be more intense with RM. We modify System (5) of our main model accordingly to accommodate for the random matching trading protocol, solve them and show the results for trading rates, trade composition, and welfare in Figure 11.

*** **Please Insert Figure 11 about here*****

Panels A, B and C of Figure 11, plot the trading rate, trade composition, and welfare, respectively. For each of these variables, we plot the difference between RM and PT as well as PBT and PT. In Panel A, we observe that in small ticks the trading rates are higher in RM compared to both PT and PBT. This is due to the dominating anticipation effect which is more pronounced with RM. However this result reverses as the half-tick increases, and leads to higher trading rates in PT. PBT does never lead to the highest trading rate.[35] Panel B displays the results for trade composition, where we can see that RM exacerbates the anticipation and queue joining effects already observed in PBT.

An analysis of the welfare differences as displayed in Panel C reveals that the relative tick size determines which of the three trading protocols yield the highest welfare. For low ticks, PT is the optimal choice for the social planner. The absence of the anticipation effect with PT leads to more LOs and relatively more investor-to-investor traders that are favorable for welfare. For "intermediate" ticks, PBT yields the highest welfare. RM then gives too much incentives to join the queue. For large ticks, RM produces the highest welfare as there are many LOs in the LOB and the anticipation effect is most pronounced. RM can be seen as a stronger form of the likelihood of queue jumping leading to more investor-to-investor trades. The two vertical lines on Panel B indicate the cutoff values for the tick regions for which each trading protocol is optimal; PT for "small", PBT for "intermediate" and RM for "wide" ticks.

[35] Our theoretical findings are consistent with the empirical findings of Lepone and Yang (2015). They exploit the introduction of pro-rata allocation to Euribor futures contracts on NYSE Liffe and find that trading volume increases substantially after the event. The Euribor futures contracts constitute a tight market.

9 Conclusion

Priority rules determine how investors reveal their trading intentions in financial markets. In this paper, we compare the impact on market quality and investor welfare of two different allocation rules that are commonly observed: price-time priority (PT) and order preferencing as modeled through price-broker-time priority (PBT). PBT for example is included in the trading protocol of Canadian and Nordic markets. While in the U.S., PT applies at an individual platform, order preferencing towards a specific platform introduces a secondary priority rule in between price and time when multiple platforms coexist. Order preferencing by brokers towards a platform when multiple markets coexist, could stem from (volume-based) maker-taker pricing, agency problems between broker and investor, or brokers having an ownership stake in a platform.

We develop a dynamic microstructure model where investors can trade either by submitting a limit order, a market order against standing limit orders, or in the absence of them trade against a dealer-specialist. Similar to Parlour (1998), the central intuition of our paper is that each trader knows that her order affects the order placement strategies of those who follow. We add to this by showing that (i) priority rules *within* or *across* markets determine the actions of subsequent investors, and (ii) rational investors anticipate subsequent investors' behavior and therefore behave differently depending upon priority rules.

Our model generates interesting insights for market quality and investor welfare. Two forces drive the results when comparing PT to PBT. First, with PBT, the *anticipation effect* induces investors to submit more often market orders when the limit order book on their side is shallow: they anticipate that a limit order could be queue jumped by later arriving investors using other brokers. Second, with PBT, investors are more inclined to *join the queue* as they may jump the queue later on.

The magnitude of these two forces differs depending upon the relative tick size, i.e., the tick compared to the heterogeneity in investor valuations. With small relative ticks, the anticipation effect is important as submitting market orders is not very costly. Trading rates (fill rates) of limit orders then are higher (lower) with PBT than with PT.

The opposite holds for wide ticks where the queue-joining effect dominates. Investor welfare is maximized with PT for small ticks and with PBT for large ticks.

We further show that each broker when given the choice between PT and PBT has a dominant strategy to implement PBT when maximizing their traders' welfare. When the relative tick is small, brokers end up in a prisonner's dilemma as they would be better off if they could commit to PT. Regulators prefer PT when the relative tick is small and PBT for wide ticks.

Systematic patterns in order flow occur and differ between PT and PBT. With PBT, it is more likely to see two consecutive limit orders or two consecutive market orders at the same side of the book than with PT. Consecutive limit orders at the same side of the book through the same (different) broker are less (more) likely under PBT than under PT. We further show that order flow on one platform depends upon previous flow at the other platform when markets are fragmented. In particular, limit orders on one platform are more likely to be followed by limit orders on the other platform than on the same platform.

Priority rules have broader implications on whether trading takes place on-exchange or off-exchange. Our model predicts more off-exchange trading under PT if brokers circumvent time priority by matching arriving market orders against their clients' standing limit orders deeper in the book. This is a novel result as it identifies one of the driving forces which push traders away from on-exchange trading.

References

Anand, Amber, Mehrdad Samadi, Jonathan Sokobin, and Kumar Venkataraman, 2019, Institutional order handling and broker-affiliated trading venues, *Working Paper* 1–44.

Angel, James J., Lawrence E. Harris, and Chester S. Spatt, 2011, Equity trading in the 21st century, *Quarterly Journal of Finance* 1, 1–53.

Aspris, Angelo, Sean Foley, Drew Harris, and Peter O'Neill, 2015, Time pro-rata matching: Evidence of a change in LIFFE Stir futures, *Journal of Futures Markets* 35, 522–541.

Battalio, Robert, Shane A. Corwin, and Robert Jennings, 2016, Can brokers have it all? On the relation between make-take fees and limit order execution quality, *Journal of Finance* 71, 2193–2238.

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

Biais, Bruno, Thierry Foucault, and Sophie Moinas, 2015, Equilibrium fast trading, *Journal of Financial Economics* 116, 292–313.

Biais, Bruno, Pierre Hillion, and Chester Spatt, 1995, An empirical analysis of the limit order book and the order flow in the Paris bourse, *Journal of Finance* 50, 1655–1689.

Bloomfield, Robert, and Maureen O'Hara, 1998, Does order preferencing matter?, *Journal of Financial Economics* 50, 3–37.

Buti, Sabrina, Francesco Consonni, Barbara Rindi, Wen Yanji, and Ingrid M Werner, 2015, Sub-penny and queue-jumping, *Working Paper* 1–53.

Buti, Sabrina, and Barbara Rindi, 2013, Undisclosed orders and optimal submission strategies in a limit order market, *Journal of Financial Economics* 109, 797–812.

Buti, Sabrina, Barbara Rindi, and Ingrid M. Werner, 2017, Dark pool trading strategies, market quality and welfare, *Journal of Financial Economics* 124, 244–265.

Chao, Yong, Chen Yao, and Mao Ye, 2017, Discrete pricing and market fragmentation: A tale of two-sided markets, *American Economic Review* 107, 196–199.

Chung, Kee H., Chairat Chuwonganant, and D. Timothy McCormick, 2004, Order preferencing and market quality on NASDAQ before and after decimalization, *Journal of Financial Economics* 71, 581–612.

Cimon, David A., 2018, Broker routing decisions in limit order markets, *Working Paper* 1–42.

Colliard, Jean Edouard, and Thierry Foucault, 2012, Trading fees and efficiency in limit order markets, *Review of Financial Studies* 25, 3389–3421.

Dahlström, Petter, Björn Hagströmer, and Lars L Nordén, 2018, Determinants of limit order cancellations, *Working Paper* 1–68.

De Jong, Frank, and Barbara Rindi, 2009, The microstructure of financial markets, Cambridge University Press.

De Winne, Rudy, and Catherine D'Hondt, 2007, Hide-and-seek in the market: Placing and detecting hidden orders, *Review of Finance* 11, 663–692.

Degryse, Hans, Mark Van Achter, and Gunther Wuyts, 2020, Plumbing of securities markets: The impact of post-trade fees on trading and welfare, *forthcoming Management Science* .

Degryse, Hans, Frank de Jong, and Vincent Van Kervel, 2015, The impact of dark trading and visible fragmentation on market quality, *Review of Finance* 19, 1587–1622.

Degryse, Hans, Mark Van Achter, and Gunther Wuyts, 2009, Dynamic order submission strategies with competition between a dealer market and a crossing network, *Journal of Financial Economics* 91, 319–338.

Duffie, Darrell, Nicolae Garleanu, and Lasse Heje Pedersen, 2005, Over-the-counter markets, *Econometrica* 73, 1815–1847.

Duffie, Darrell, Nicolae Garleanu, and Lasse Heje Pedersen, 2007, Valuation in over-the-counter markets, *Review of Financial Studies* 20, 1865–1900.

Dugast, Jerome, 2018, Unscheduled news and market dynamics, *Journal of Finance* 73, 2537–2586.

EMA, 2017, SEC tick size pilot, Equity Markets Association, https://www.nyse.com/publicdocs/ EMA_Report_SEC_Tick_Pilot_Data +Preliminary_Findings.pdf.

ESRB., 2016, Market liquidity and dividends, European system of financial supervision.

Field, Jonathan, and Jeremy Large, 2008, Pro-Rata matching and one-tick futures markets, *CFS, Working Paper* 1–29.

Foley, Sean, Elvis Jarnecic, and Anqi Liu, 2019, Forming an orderly line - How queue-jumping drives excessive fragmentation, *Working Paper* 1–31.

Foucault, Thierry, 1999, Order flow composition and trading costs in a dynamic limit order market, *Journal of Financial Markets* 2, 99–134.

Foucault, Thierry, Ohad Kadan, and Eugene Kandel, 2005, Limit order book as a market for liquidity, *Review of Financial Studies* 18, 1171–1217.

Foucault, Thierry, Ohad Kadan, and Eugene Kandel, 2013a, Liquidity cycles and make/take fees in electronic markets, *Journal of Finance* 68, 299–341.

Foucault, Thierry, and Albert J. Menkveld, 2008, Competition for order flow and smart order routing systems, *Journal of Finance* 63, 119–158.

Foucault, Thierry, Marco Pagano, and Alisa Roell, 2013b, Market liquidity: Theory, evidence, and policy, Oxford University Press.

Goettler, Ronald L., Christine A. Parlour, and Uday Rajan, 2005, Equilibrium in a dynamic limit order book market, *Journal of Finance* 60, 2149–2193.

Goettler, Ronald L., Christine A. Parlour, and Uday Rajan, 2009, Informed traders and limit order markets, *Journal of Financial Economics* 93, 67–87.

Handa, Puneet, Robert Schwartz, and Ashish Tiwari, 2003, Quote setting and price formation in an order driven market, *Journal of Financial Markets* 6, 461–489.

Harris, Larry, 2013, Maker - taker pricing effects on market quotations, *Working Paper* 1–50.

Haynes, Richard, and Esen Onur, 2019, Precedence rules in matching algorithms, *Journal of Commodity Markets, forthcoming* 1–11.

Hendershott, Terrence, and Haim Mendelson, 2000, Crossing networks and dealer markets: Competition and performance, *Journal of Finance* 55, 2071–2115.

Hendershott, Terrence, and Albert J. Menkveld, 2014, Price pressures, *Journal of Financial Economics* 114, 405–423.

Hoffmann, Peter, 2014, A dynamic limit order market with fast and slow traders, *Journal of Financial Economics* 113, 156–169.

Hollifield, Burton, Robert A. Miller, Patrik Vilhelm Sandas, and Joshua Slive, 2006, Estimating the gains from trade in limit order markets, *Journal of Finance* 61, 2753–2804.

IOSCO, 2017, Annual report, 1-90.

Kwan, Amy, Ronald Masulis, and Thomas H. McInish, 2015, Trading rules, competition for order flow and market fragmentation, *Journal of Financial Economics* 115, 330–348.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1335.

Lepone, Andrew, and Jin Young Yang, 2012, The impact of a pro-rata algorithm on liquidity: Evidence from the NYSE Liffe, *The Journal of Futures Markets* 32, 660–682.

Li, Sida, Xin Wang, and Mao Ye, 2019, Who provides liquidity and when?, *Working Paper* 1–64.

Malinova, Katya, and Andreas Park, 2015, Subsidizing liquidity: The impact of make/take fees on market quality, *Journal of Finance* 70, 509–536.

Michaels, Dave, 2018, Curtains for experiment meant to boost trading in small stocks, WSJ, https://www.wsj.com/articles /curtains-for-experiment-meant-to-boost-trading-in-small-stocks-1523387965.

O'Hara, Maureen, Gideon Saar, and Zhuo Zhong, 2019, Relative tick size and the trading environment, *The Review of Asset Pricing Studies* 9, 47–90.

Pagnotta, Emiliano S, 2010, Information and liquidity trading at optimal frequencies, *Working Paper* 1–53.

Parlour, Christine A., 1998, Price dynamics in limit order markets, *Review of Financial Studies* 11, 789–816.

Parlour, Christine A., and Duane J. Seppi, 2003, Liquidity-based competition for order flow, *Review of Financial Studies* 16, 301–343.

Pearson, Phil, and Colleen Ruane, 2017, Tick pilot update, ITG, http://www.itg.com/assets/ITG-Tick-Pilot-Update-2017.pdf.

Ranaldo, Angelo, 2004, Aggressiveness in limit order book markets, *Journal of Financial Markets* 7, 53–74.

SEC, 1997, Report on the practice of preferencing, U.S. Securities and Exchange Commission, www.sec.gov/news/studies/prefrep.htm.

SEC, 2005, Regulation NMS. Release No. 34-51808. File No. s7-10-04, U.S. Securities and Exchange Commission.

Seppi, Duane J., 1997, Liquidity provision with limit orders and strategic specialist, *Review of Financial Studies* 10, 103–150.

Spatt, Chester S., 2019, Is equity market exchange structure anti-competitive?, *Working Paper* 1–37.

Van Kervel, Vincent, 2015, Competition for order flow with fast and slow traders, *Review of Financial Studies* 28, 2094–2127.

Yueshen, Bart Zhou, 2014, Queuing uncertainty in limit order market, *Working Paper* 1–56.

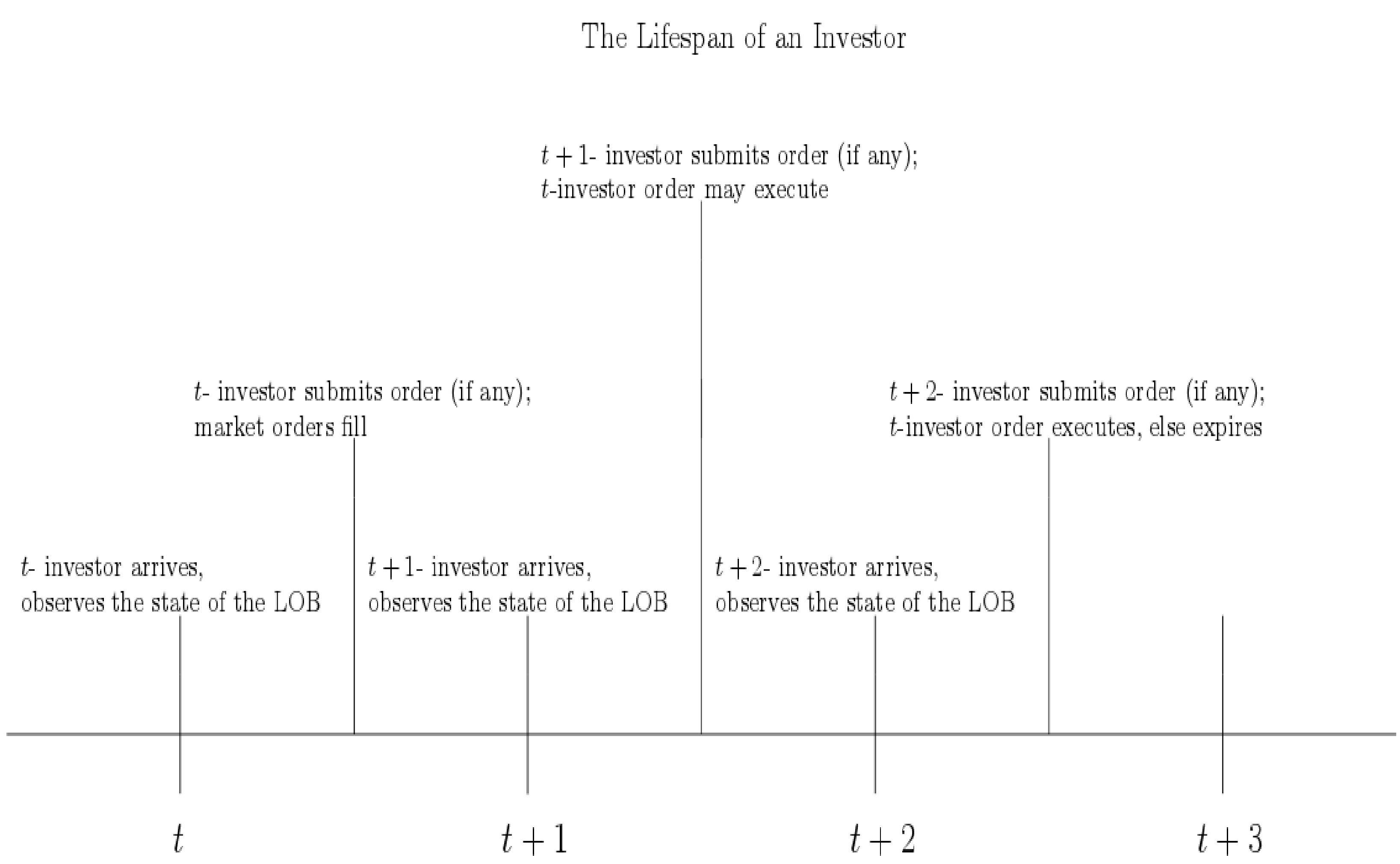

Figure 1: Representation of the lifespan and timing of actions for an investor.

This figure illustrates the lifespan of the arriving investor and the timing of her actions. At time t, the trader arrives, observes the LOB and forms her actions. A limit order remains in the book for two periods, and if not filled within that time span, it expires. Her execution probability depends on the secondary priority rules.

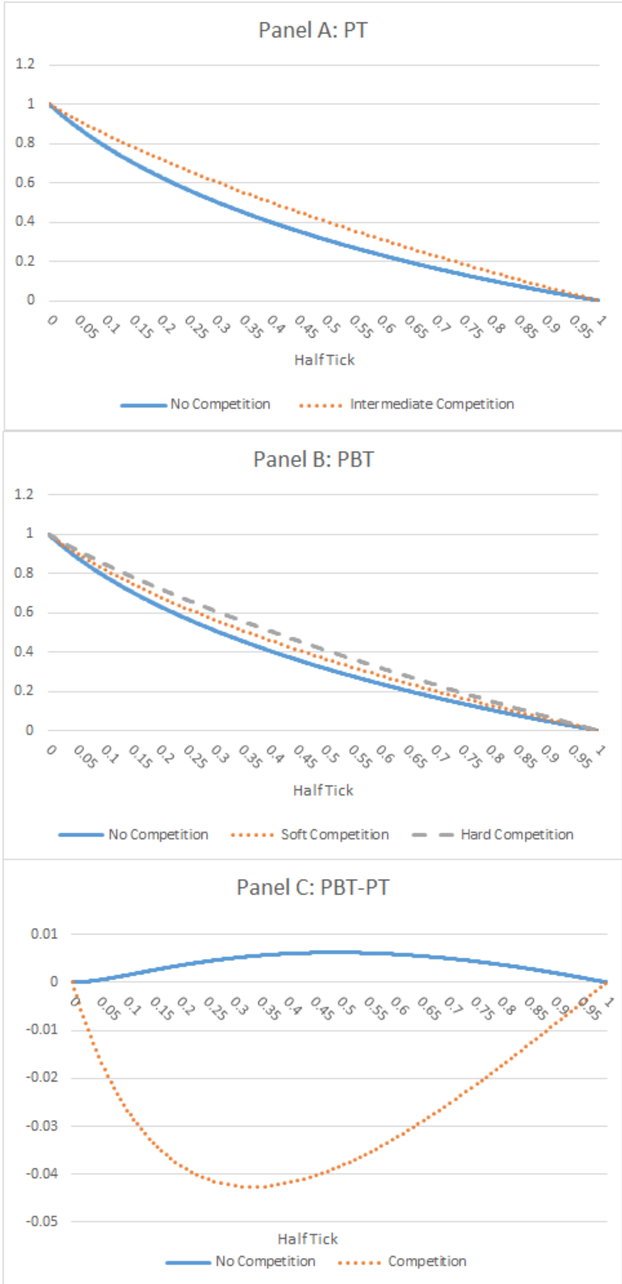

Figure 2: Cut-off values for PT and PBT

This figure illustrates the cut-off values of a seller facing different levels of competition for all tick sizes for a traded asset where we assume $V = 1$. Panel A (B) identifies the cut-off values under PT (PBT) for different levels of competition, i.e., no competition and intermediate competition for PT and no-competition and 'soft' and 'hard' competition for PBT. In Panel C, their difference is depicted. For the difference under competition, in PBT we consider the average of the cut-off values between 'soft' and 'tough' competition.

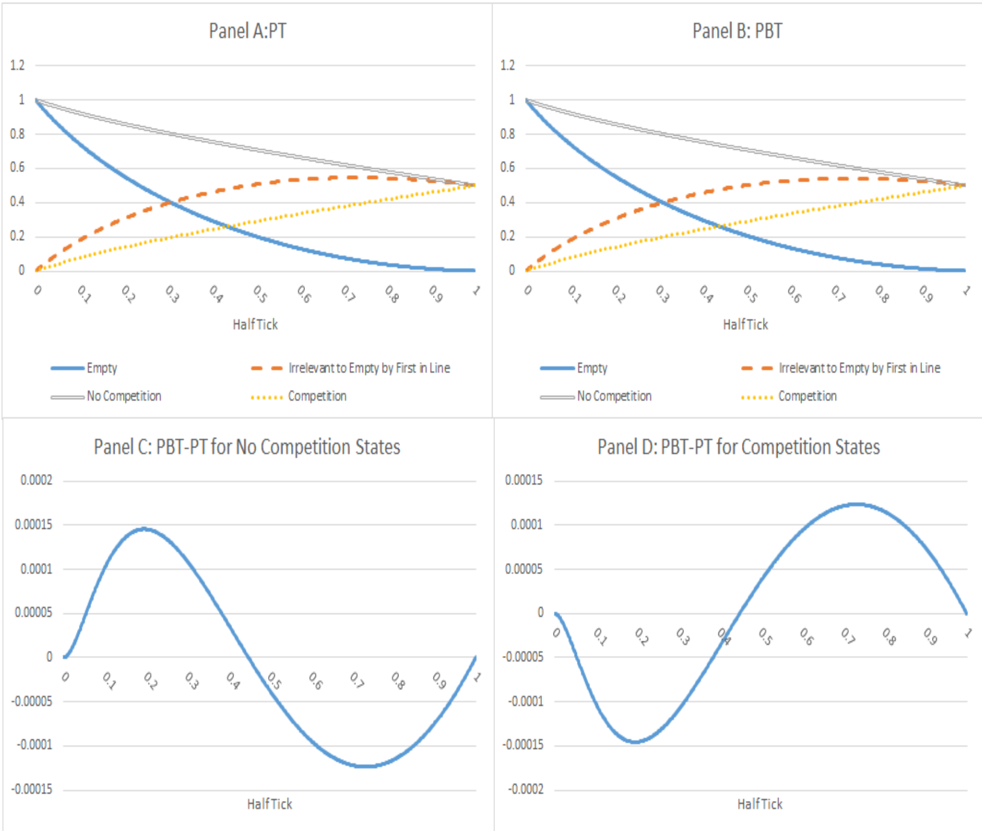

Figure 3: Consolidated Markov Steady States Probabilities for PT and PBT

This figure illustrates the consolidated Markov state probabilities derived as the solution of the matrix equation

$$\rho_i = \rho_i \mathcal{M}_i,$$

for all tick sizes and various levels of competition, for both PT and PBT(Panels A and B respectively), as well as the difference PBT-PT for no-competition and competition states (Panels C and D). We denote by i the trading protocol, PT or PBT and $\mathcal{M}_i$ denotes the transition matrix of the Markov chain under the protocol i and ρ_i represents the vector of all states. '*Empty*' stands for the steady state probability of observing a completely empty book and '*Irrelevant to Empty by First in Line*' graphs the probability that an arriving seller faces an irrelevant to an empty state. '*No competition*' refers to their sum, and '*Competition*' illustrates the consolidated probabilities that would place a trader upon arrival, to a state in which she would face competition, intermediate for PT (Panel A) and combined soft and tough for PBT (Panel B). Panel C (D) shows the difference in probability between PBT and PT for the states that put the arriving trader under no competition (competition). We assume $V = 1$.

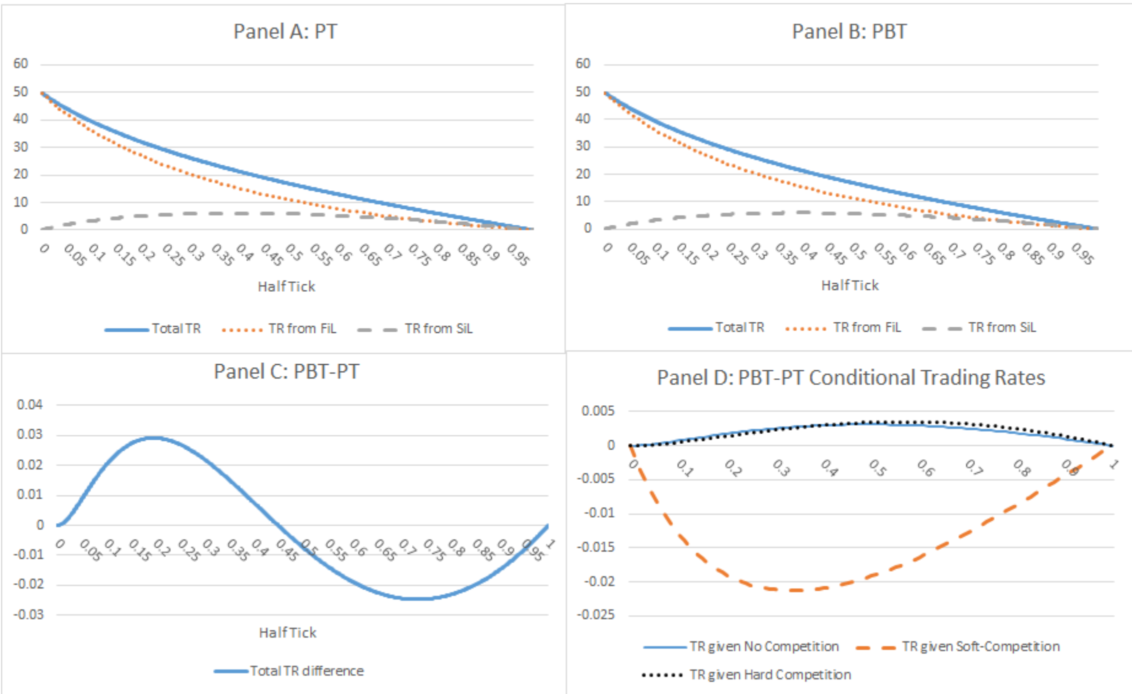

Figure 4: Trading Rates for PT and PBT

This figure illustrates the trading rates in PT and PBT as a function of the half-tick following the calculations of Section 3.1.1. In Panel A (B), the total trading rate, and the trading rate by a seller that faces no-competition and competition in PT (PBT), are graphed. Panel C illustrates the difference in the total TR between PBT and PT for all tick sizes. In Panel D, we plot the difference in the conditional trading rates in PBT and PT depending upon the competition that the arriving trader faces. For the computations under competition in Panel D for PT, intermediate competition has been used. We assume $V = 1$.

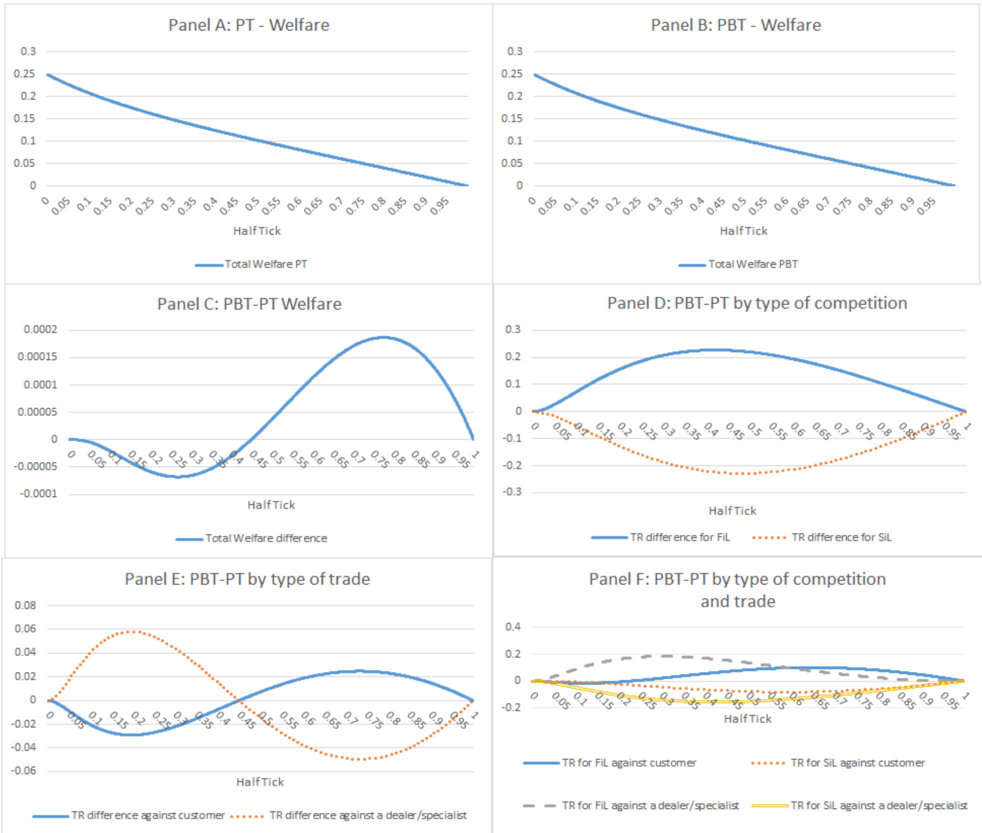

Figure 5: Investor Welfare and trade composition

This figure illustrates the difference in investor welfare between PT and PBT, as well as differences in trading rates for various groups of traders, for all tick sizes. For the computations we follow the discussion presented in Section 3.2. Panel A (B), plots the overall generated investor welfare for PT (PBT). Panel C graphs the difference in overall welfare between PBT and PT. Panel D illustrates the difference in trading rates by traders facing no-competition and competition. In Panel E, we plot the difference in trading rates between PBT and PT for customer-customer trades and customer-dealer/specialist trades. Panel F further decomposes customer-customer trades and customer-dealer/specialist trades initiated by a first or second-in-line arriving agent, i.e., by the degree of competition. We assume $V = 1$.

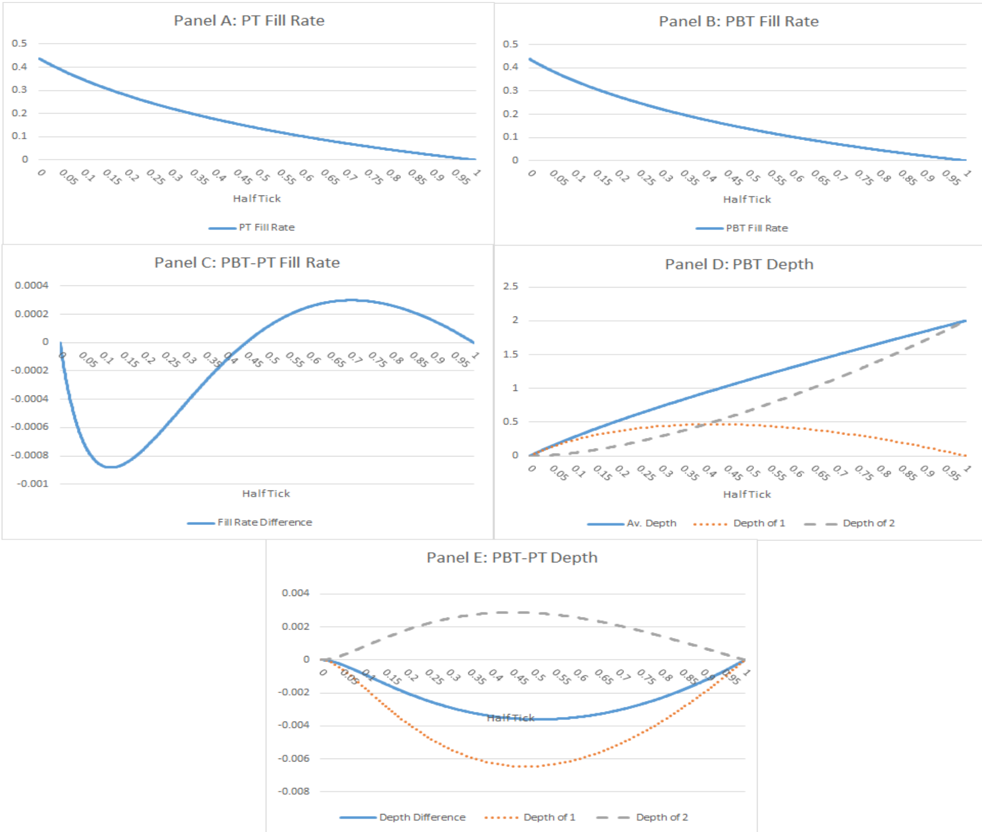

Figure 6: Depth and Fill Rate under PBT and PT

The figure illustrates the fill rate of limit orders and depth as a function of the tick size. Panel A (B) shows the fill rates in PT (PBT), while in Panel C we chart the difference in the fill rate of a sell LO between the two trading protocols. Panel D graphs the average depth, the (unconditional) 'depth of 1' and 'depth of 2' in PBT as a function of tick size. Panel E, plots the difference between PBT and PT for the same variables as in Panel D. We assume $V = 1$.

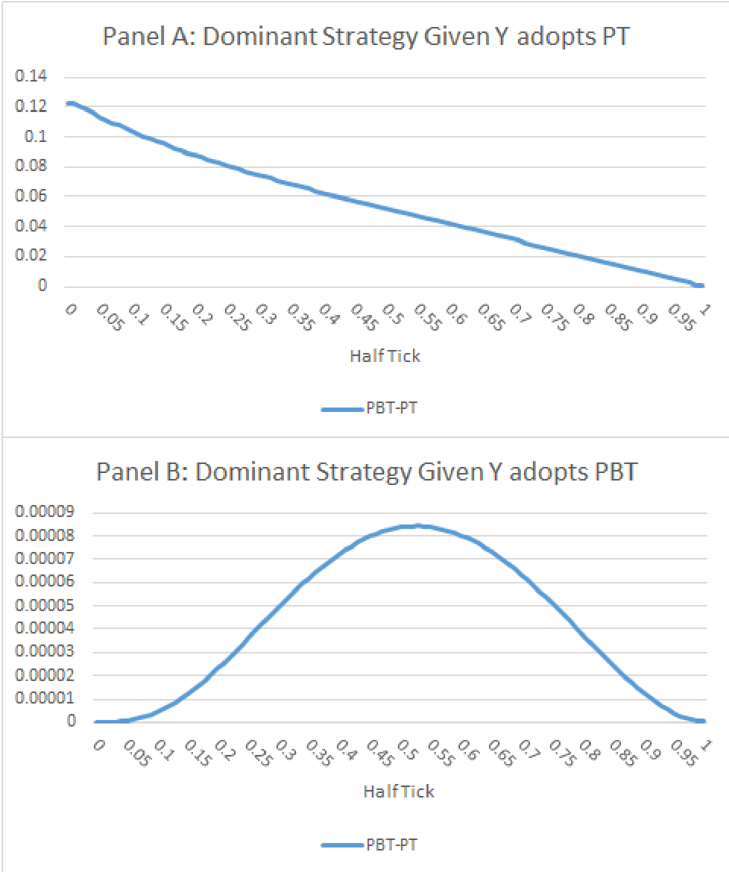

Figure 7: Endogenous adoption of priority rules: difference in the pay-off matrix

The figure illustrates the difference in payoffs for broker X, assuming broker Y adopts PT (Panel A) or PBT (Panel B). For a $2-by-2$ symmetrical pay-off matrix the graphs show that for all tick sizes the unique Nash equilibrium is found when both of brokers implement PBT. We assume $V = 1$. For details please see Table D2, in Internet Appendix D.

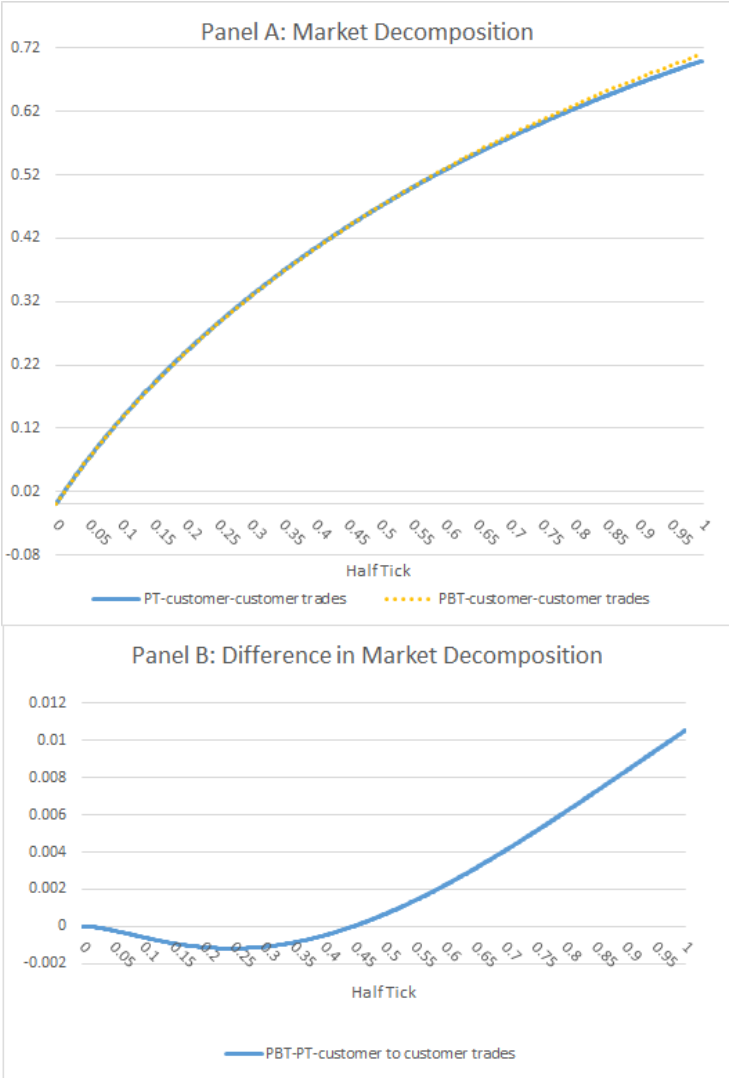

Figure 8: Market share of customer-customer trades for PBT and PT

The figure shows the market shares of customer-customer trades (Panel A) for both PT (solid line) and PBT (dotted line) as well their difference (Panel B). In Panel A, for each trading protocol, the complementary market share identifies the customer-dealer/specialist trades. The half tick size for which we observe a change from lower market share to higher, for PBT relative to PT, corresponds to $\tilde{\Delta}/2$, (see, Section 3.1.2).

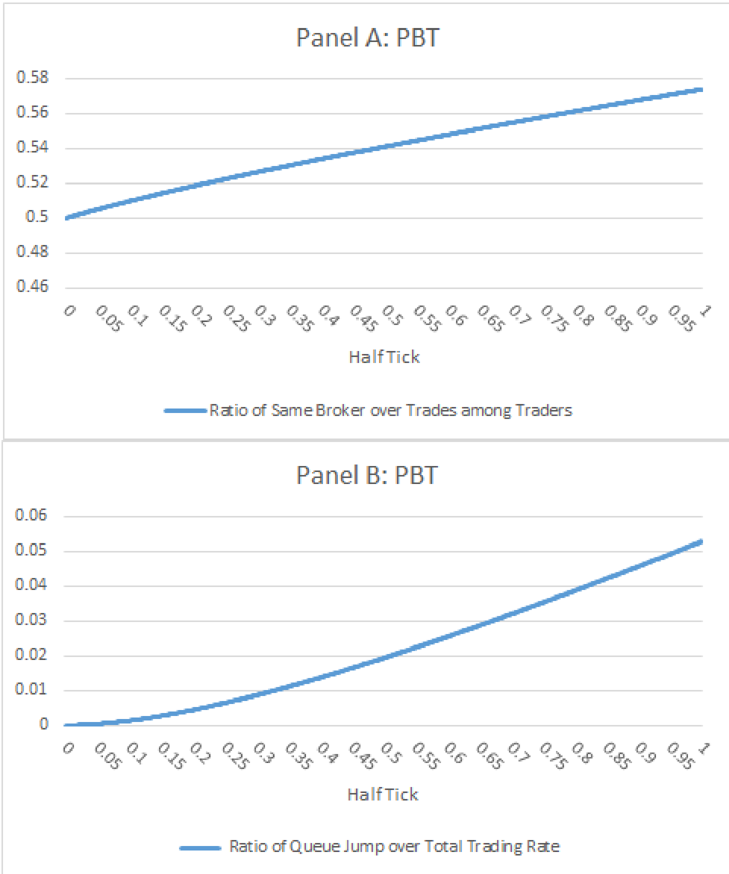

Figure 9: 'Same Broker' matching, queue jumping, and off-exchange reporting

The figure illustrates how PBT affects the degree of 'same-broker' trades and the importance of queue jumping in the total trading rate for all tick sizes for a traded asset with fundamental value V normalized to one. Panel A reports the ratio of matched customer-customer trades between traders having the same broker over the total amount of trades as a function of the half-tick. In Panel B we plot the ratio of actual queue jumping in PBT over the total trading volume as a function of the half-tick. Notice that the latter is identical to the degree of off-exchange trading when brokers would circumvent PT by reporting off exchange.

Figure 10: Trading rates and investor welfare in the absence of dealer-specialists

This figure plots trading rates and investor welfare for all tick sizes assuming $V = 1$, in the absence of dealer-specialists. Trading rates and welfare are obtained by a suitable modification of our main model presented in Section A. Panel A charts the difference in trading rates between PBT and PT while in Panel B we plot the difference in welfare as a function of the half-tick.

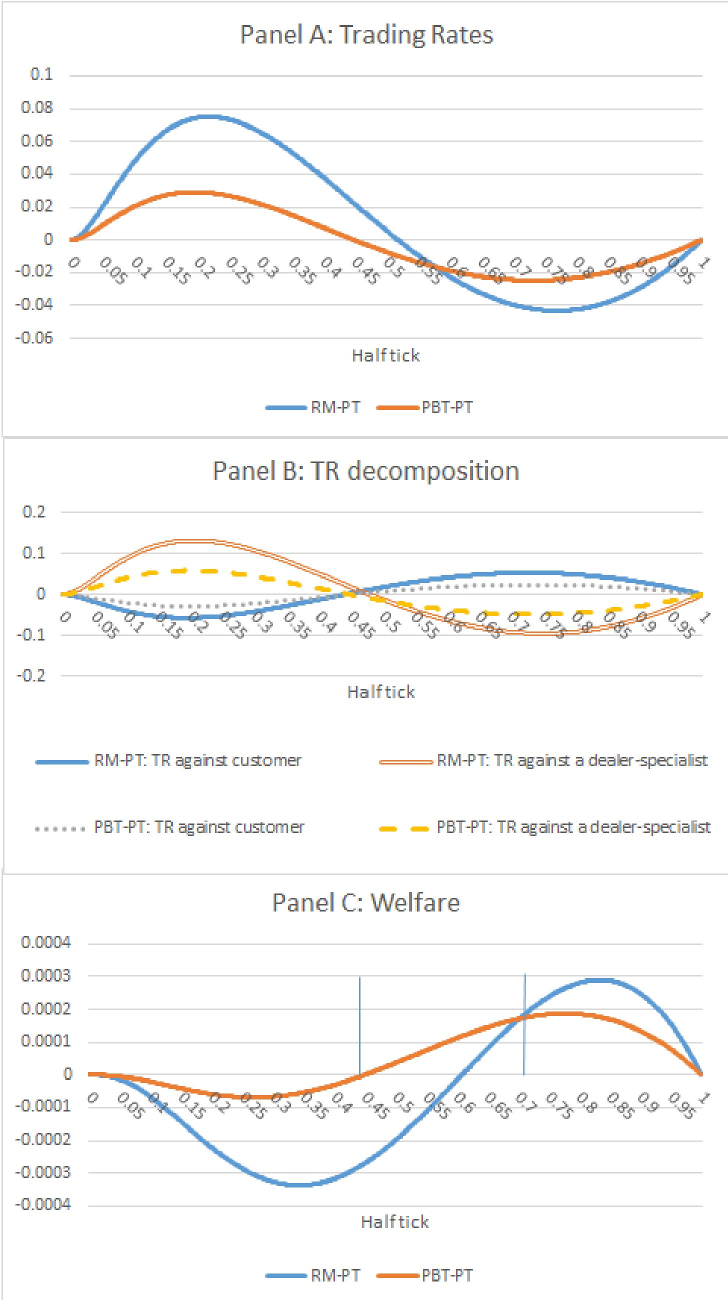

Figure 11: Random matching: trading rates and investor welfare

This figure plots differences in total and decomposed by type, trading rates (Panel A and B), and investor welfare (Panel C) for all tick sizes assuming $V = 1$. Panel A plots the difference in trading rate between RM and PT as well as PBT and PT. In Panel B, we plot the difference in likelihood of customer-customer trades and customer-dealer/specialist trades, between RM and PT as well as PBT and PT. Panel C graphs the difference in generated welfare by the trading protocols. The two vertical lines indicate the tick regions which identify the optimal trading protocol; PT for “small”, PBT for “intermediate” and RM for “wide” ticks.

Internet Appendix

A Model's Equations

In order to depict the solution to the decision problem of the arrivnig to the market seller which in term identifies the equilibrium of the model, we need to consider both broker affiliations. The seller needs to solve the following system.

$$\Sigma = \begin{Bmatrix} \Sigma_x \\ \Sigma_y \end{Bmatrix}, \tag{2}$$

where Σ_k defines a set of equations that makes the k-seller, $k \in \{x, y\}$, indifferent between a market and a limit order considering the state of the book upon her arrival. Since brokers have equal market shares, we obtain that $\Sigma_x = \Sigma_y$ and hence we can focus to an x-seller.[36] Its solution generates three cut-off values for the arriving x-seller.

If the arriving trader opts for a limit order, then the execution probability depends on the flow of future traders.[37] For every time period, the arriving trader needs to account for all potential flow of traders when calculating her gain from submitting a limit order. At the same time, she needs to account also for the particular state of the book at her side, at the time of her arrival. Let $G_b = G_b(A, B, X, Y)$ denote the probability of execution of a submitted sell limit order, if the next arriving trader is a buyer taking the action of submitting a limit order, or decline trading and by $G_s = G_s(A, B, X, Y)$ the execution probability of a submitted sell limit order, if the next arriving trader is a seller who either submits a market order, a limit order or refrain from trading.

G_b and G_s are defined as follows:

[36] Because of size symmetry, a seller affiliated to broker X has the same cut-off values with a seller affiliated with broker Y.

[37] For example, for a sell limit order at time t on an empty book, one potential flow would be that at the next period a buyer arrives and submits a limit order to buy. That means that the seller in order to obtain an execution would need at $t + 2$ the arrival of buyer who would submit a market order.

$$
\begin{aligned}
G_b =& \Big[\frac{x}{2}P_{t+1}((B-Z_0)/V \leq b \leq b_x^B(0,-1^x))\big(\frac{x}{2}P_{t+2}(b \geq b_x^B(1^x,0))+ \\
&\frac{y}{2}P_{t+2}(b \geq b_y^B(1^x,0))\big) + \frac{y}{2}P_{t+1}((B-Z_0)/V \leq b \leq b_y^B(0,-1^x)) \\
&\big(\frac{x}{2}P_{t+2}(b \geq b_x^B(1^y,0)) + \frac{y}{2}P_{t+2}(b \geq b_y^B(1^y,0))\big) + \frac{1}{2}P_{t+1}(b \leq (B-Z_0)/V)) \\
&\big(\frac{x}{2}P_{t+2}(b \geq b_x^B(0,0)) + \frac{y}{2}P_{t+2}(b \leq b_y^B(0,0))\big)\Big]
\end{aligned}
\tag{3}
$$

and equivalently

$$
\begin{aligned}
G_s =& \frac{x}{2}P_{t+1}(b \leq b_x^S(0,-1^x)) \\
&\big(\frac{x}{2}P_{t+2}(b \geq b_x^B(0,0)) + \frac{y}{2}P_{t+2}(b \geq b_y^B(0,0))\big)+ \\
&\frac{y}{2}P_{t+1}(b \leq b_y^S(0,-1^x))\big(\frac{x}{2}P_{t+2}(b \geq b_x^B(0,0)) + \frac{y}{2}P_{t+2}(b \geq b_y^B(0,0))\big)+ \\
&\frac{x}{2}P_{t+1}(b_x^S(0,-1^x) \leq b \leq (A-Z_0)/V)\big(\frac{x}{2}P_{t+2}(b \geq b_x^B(0,-1^x))+ \\
&\frac{y}{2}P_{t+2}(b \geq b_y^B(0,-1^x))\big) + \frac{y}{2}P_{t+1}(b_x^S(0,-1^x) \leq b \leq (A-Z_0)/V) \\
&\big(\frac{x}{2}P_{t+2}(b \geq b_x^B(0,-1^x)) + 0\big) + \frac{1}{2}P_{t+1}(b \geq (A-Z_0)/V)\big(\frac{x}{2}P_{t+2}(b \geq b_x^B(0,0)) \\
&+ \frac{y}{2}P_{t+2}(b \geq b_y^B(0,0))\big)\Big]
\end{aligned}
\tag{4}
$$

In the definitions above, both x and y are equal to 0.5, but are stated as such in order to identify the desired broker affiliation. Notice that G_b and G_s do not depend on the state that the seller arrives at t. This is a direct consequence of the two period exogenous cancellation rule. Therefore, each trader needs to account for the same values of G_b and G_s. What distinguishes the execution probability of a sell limit order related to the state of the book of the arriving seller, is the action of the subsequent buyer, should she arrive at $t+1$, and whether she would submit a market order. In this case, the state of the book could determine an immediate execution, a preferential execution or neither. Given this discussion, we are ready to define the system Σ_x that the arriving x-seller needs to solve.

$$
\begin{cases}
(B-(Z_0+b_x^S(0,0)V)) = & \big(\frac{x}{2}P_{t+1}(b\geq b_x^B(0,-1^x))+\frac{y}{2}P_{t+1}(b\geq b_x^B(0,-1^x))\big) \\
& (A-(Z_0+b_x^S(0,0)V))+(G_b+G_s)(A-(Z_0+b_x^S(0,0)V)) \\
(B-(Z_0+b_x^S(0,-1^y)V)) = & \frac{x}{2}P_{t+1}(b\geq b_x^B(0,-1^x))(A-(Z_0+b_x^S(0,0)V)) \\
& +\frac{y}{2}P_{t+1}(b\geq b_y^B(0,-1^x)) \\
& \big(\frac{x}{2}P_{t+2}(b\geq b_x^B(0,0))+\frac{y}{2}P_{t+2}(b\geq b_y^B(0,0))\big) \\
& (A-(Z_0+b_x^S(0,-1^y)V)) \\
& +(G_b+G_s)(A-(Z_0+b_x^S(0,-1^y)V)) \\
(B-(Z_0+b_x^S(0,-1^x)V)) = & \big(\frac{x}{2}P_{t+1}(b\geq b_x^B(0,-1^x))+\frac{y}{2}P_{t+1}(b\geq b_y^B(0,-1^x))\big) \\
& \big(\frac{x}{2}P_{t+2}(b\geq b_x^B(0,0))+\frac{y}{2}P_{t+2}(b\geq b_x^B(0,0))\big) \\
& (A-(Z_0+b_x^S(0,-1^x)V)) \\
& +(G_b+G_s)(A-(Z_0+b_x^S(0,-1^x)V))
\end{cases}
\tag{5}
$$

In System 5, both x and y are equal to 0.5, but we include them as such in order to demonstrate the required trader flow and broker affiliation needed to obtain execution. In each of the equations, the LHS defines the gains of the seller from submitting a market order depending on the state of the book that she faces. The RHS reports the gains from a limit order to sell. In the first equation, we notice that the seller may get execution at the next period, since she faces no competition, in comparison to the last equation where she needs to stand in the line for two periods. The equation in the middle describes the difference in the execution probability which is due to PBT. The seller, even second in line, can get an execution the following period through her broker flow by the use of preferential services. Under PT, the arriving seller solves a similar but simplified version of System 5.

The arriving trader, can perform one of three actions. Trade via a market order (at the dealer market or LOB), submit a limit order or refrain from trading. Her decision depends on the state of the book and her willingness to trade b. The critical values which identify the actions of a trader are determined by the solution of System 5. One of the main implications of our model is that under PBT, a trader faces three different cut-off values in relation to the submission of a LO, while in PT these reduce to two. The reason being that under PBT a trader's decision is also affected by the structure of the queue, and in particular she distinguishes between tough and soft competition. In PT, a trader facing a non empty book on her side, always faces intermediate competition, regardless the structure of the queue. The book is formed based on traders' actions. Since the orders do not stay in the book indefinitely and are canceled exogenously, the states of the book are finite and define a Markov chain. Let $\mathcal{M}$ denote the transition matrix of the Markov chain, i.e. the probability that a trader transits from one state of the book to another. We note that these probabilities depend on the random sequence of the arriving traders and from the willingness to trade. Following Foucault et al., (2005), we derive the stationary probabilities ρ of the system. The stationary distribution shows the probability that the arriving trader faces a specific state of the book independent to the exact time of the arrival. In contrast to Degryse et al., (2020), we identify the likelihood that an arriving trader faces a specific state of the book and not performing a certain action.[38] Table D1 in Internet Appendix D depicts them in detail. The steady state equilibrium is defined as the left eigenvector of the transition matrix, which is given as the solution of the matrix equation

$$\rho = \rho\mathcal{M}.$$

We need to notice that PT in comparison to PBT differ in both the number of the relevant states as well as the transition matrix for a given trader. In PT, for example, even though being second-in-line after a trader of the same or opposite affiliation defines two distinct states, the arriving trader treats them as the same as it is immaterial to her. The defined Markov chain, in both models is irreducible and aperiodic and thus we obtain unique distribution for both systems (see, Internet Appendix C).

[38]Essentially these are equivalent. From the steady state distribution of the book and the distribution of β, we can derive the steady distribution of the actions.

B Proofs and results

In the remaining of the section, we report the proofs, obtained analytical and numerical (i.e., proof via exhaustion) of the propositions stated in the main body of the text.

Proposition 1: The first part of the claim is obvious. We will show that the trader's decision does not depend on the state of the book on the opposite side. Assume an x-seller arriving to the market. We need to show that her cut-off value depends on the state of the book on her side only. Let us analyze only one potential case, since all the rest follow applying similar arguments. We compare the following states of the book, $(0, 0)$ and $(1^x, 0)$, i.e., a completely empty book and a book that has one limit order standing at the bid. Let $b_x^S(0, 0)$ and $b_x^S(1^x, 0)$ be the two corresponding cut-off values. We need to show that $b_x^S(0, 0) = b_x^S(1^x, 0)$. We denote $P(LO(s_i))$ the execution probability of the limit order when the state of the book is s_i. We notice

$$P(LO(0, 0)) = P(LO(1^x, 0)).$$

To see that, assume that the x-seller at t, submits a limit order and let the arriving trader $t + 1$ be seller. If the book at t was $b_x^S(0, 0)$ then this trader faces the state $b_x^S(0, -1^x)$. If the book was $b_x^S(1^x, 0)$ then the trader faces $b_x^S(1^{2,x}, -1^x)$ which is equivalent to $b_x^S(0, -1^x)$ concerning the actions of the arriving trader. The same reasoning holds if at $t+1$ a buyer arrives. Thus the decision of the trader at $t+1$ does not depend on the state of the book on the opposite side that the x- seller faces at t. A similar reasoning holds for the subsequent period and hence, the execution probability of the submitted limit order at t, in both cases is equal, i.e., $P(LO(0, 0)) = P(LO(1^x, 0))$. Since the outside option of a trader on the two states of the book is the same, it results that her behavior, will be identical i.e. $b_x^S(0, 0) = b_x^S(1^x, 0)$. If she had different cut-off values facing these two different states of the book, that would be reflected in a difference in the execution probability of the limit orders. ■

The proof of **Proposition 2** results from the definition of the indifference equation between a MO and a LO.

Proposition 3: For (i), it is enough to observe that traders, buyers and sellers, exist with same percentages in the population. Moreover, the willingness to trade b, is a draw from a uniform distribution with support over $(0, 2)$ independent to the type of a trader.

For reasons of completeness we will sketch the proof for one state of the book and the remaining follow with similar arguments. Assume that the book is empty. We first notice that a buyer or a seller if they place a LO they have the same probability of execution when both facing an empty book. So we have the following system of indifference equations:

$$\begin{cases} (B - b_x^S(0,0)(V + Z_0)) = P(A - b_x^S(0,0)(V + Z_0)) \\ (b_x^B(0,0)(V + Z_0) - A) = P(b_x^B(0,0)(V + Z_0) - B) \end{cases}$$

Solving the system above and substituting one solution to the other we obtain $b_x^S(0,0)$ as a function of $b_x^B(0,0)$, and in particular $b_x^B(0,0) = 2 - b_x^S(0,0)$.

We have that (ii) results from (i) and the properties of cumulative distribution functions. ■

The proof of **Propositions 4** and **5** follow directly by the solution of System 5.

The proofs for **Propositions 6, 7 and 8** are provided via mathematical exhaustion, replicating our results for a wide range of fundamental values and granular tick sizes much finer than the 1 cent tick size found in most markets. For details please see, Section 2.3.

Proposition 9, follows directly from the corresponding system of indifference equations that a trader under opacity solves.

C Markov Steady State Equilibrium and Transition Matrix

As explained in the main body of the text, the evolution of the limit order book, defines a Markov chain for which the steady state equilibrium is the stationary distribution ρ

that satisfies the following matrix equation

$$\rho = \rho\mathcal{M},$$

where $\mathcal{M}$ is the transition matrix of the chain. Notice that the transition matrix remains a function of the market shares of the two brokers. The solution provides the steady state distribution, i.e., the likelihood that the arriving trader faces a specific state of the book after a sufficient number of traders have arrived to the market. The steady state distribution exists and is unique. The existence and uniqueness is ensured because the chain is irreducible and a-periodic. Irreducibility means that there is always a positive probability that from a particular state of the book, to be able to return on that state in finite period of time. A-periodicity, is usually more difficult to verify. A state j is periodic with period k, if starting from j we need a multiple of k steps to return to that state. If $k = 1$ then the state is a-periodic. However, given that our Markov chain is irreducible, in order to be a-periodic, it only needs one a-periodic state, which is given when the state of the book is empty. Thus there exists a unique vector representing the steady state probability distribution. Equivalent, this stationary distribution, is the left eigen vector of the transition matrix (Daroch and Senenta, 1965).

D Complementary Tables

Table D1: Irrelevant States For a Seller

This table depicts the irrelevant states for a seller i.e. the states that even though different, lead to the same action for the arriving seller. Column 1, provides the three distinct cases from which the arriving seller determines her action. Column 2 correspond to PT and PBT. Under PT a trader does not distinguishes broker affiliations, thus combines rows two and three.

	PT and PBT
Irrelevant states to $(0, 0)$, for a seller	$(0, -1^{2,x}), (1^{2,x}, 0), (0, -1^{2,y}),$ $(1^{2,y}, 0), (1^x, 0), (1^y, 0),$ $([1^{2,x}, 1^x], 0), ([1^{2,y}, 1^y], 0), ([1^{2,y}, 1^x], 0),$ $([1^{2,x}, 1^y], 0), (1^x, -1^{2,x}), (1^y, -1^{2,y}),$ $(1^x, -1^{2,y}), (1^y, -1^{2,x})$
Irrelevant states to $(0, -1^x)$, for a seller	$(0, [-1^x, -1^{2,x}]), (0, [-1^x, -1^{2,y}]),$ $(1^{2,x}, -1^x), (1^{2,y}, -1^x)$
Irrelevant states to $(0, -1^y)$, for a seller	$(0, [-1^y, -1^{2,y}]), (0, [-1^y, -1^{2,x}]),$ $(1^{2,x}, -1^y), (1^{2,y}, -1^y)$

Table D2: Brokers' decision to adopt PT or PBT

This table summarizes the Nash Equilibrium in a 2-by-2 symmetrical game. We denote by $G^X_{i,j}$, the profit of broker X, if he adopts the trading protocol i, while the Y broker adopts j. Equivalent definition holds for $G^Y_{i,j}$. If $G^Y_{PT,PBT} > G^Y_{PT,PT}$ and $G^Y_{PBT,PBT} > G^Y_{PBT,PT}$, then brokers are in a Nash equilibrium by selecting to offer mutually PBT. For a matrix 2-by-2 representation, see table below.

		Broker Y			
		PT		PBT	
Broker X	PT	$G^X_{PT,PT}$	$G^Y_{PT,PT}$	$G^X_{PT,PBT}$	$G^Y_{PT,PBT}$
	PBT	$G^X_{PT,PBT}$	$G^Y_{PBT,PT}$	$G^X_{PBT,PBT}$	$G^Y_{PBT,PBT}$