UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Financial Intermediation and International Economics

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Economics

by

Mariano Joaquin Palleja

2024

ABSTRACT OF THE DISSERTATION

Essays on Financial Intermediation and International Economics

by

Mariano Joaquin Palleja

Doctor of Philosophy in Economics

University of California, Los Angeles, 2024

Professor Pierre-Olivier Weill, Chair

This dissertation consists of three essays on financial intermediation and international economics. In the first two essays, I study how new regulations and technologies affect liquidity in decentralized over-the-counter (OTC) markets. These markets are defined by the lack of a centralized exchange, which forces customers to search for trading counterparties and encourages dealers to provide financial intermediation. In the first essay, I address the trade-off between trading speed and transaction costs investors face in a context where dealers face higher regulatory costs. In the second essay, I explore portfolio trading, the latest innovation in the corporate bond market –one of the biggest OTC markets–, highlighting its effect on market liquidity. In the third essay, I consider a scenario where countries issue assets with different liquidity and study its macroeconomic and asset pricing effects.

In recent years, stringent financial regulations and advancing trading technologies have reshaped over-the-counter intermediation, discouraging dealers from providing immediacy to customers using their own inventories (principal trades) in favor of a larger matchmaking activity (agency trades). The first chapter of this dissertation studies how customers optimally choose between these two trading mechanisms and the implications of this choice for market liquidity. I develop a quantitative search model where heterogeneous customers choose between immediate but expensive and delayed but less costly trades, i.e., principal and agency trades, respectively. Each customer solves this speed-cost trade-off, jointly determining her optimal mechanism, transaction costs, and trading volume. When market conditions change, customers migrate across mechanisms in pursuit of higher trading surpluses. I show that this migration is not random, thus liquidity measures change not only because of changes in market conditions but also because of a composition effect. To quantify such an effect, I structurally estimate my model and build counterfactual measures that control for migration. I replicate the major innovations seen in these markets and find that composition effects explain more than a third of the increase in principal transaction costs.

The second chapter studies a recent innovation in the corporate bond market: portfolio trading. In contrast to sequential trading, this new protocol allows customers to trade a list of bonds as a single security. I show that these trading features have significant consequences on market liquidity. Particularly, I present novel evidence of asymmetrical transaction costs: compared to sequential trading, portfolio trading is less expensive when customers buy bonds and more expensive when they sell them. I find that dealers' balance sheet costs and portfolios' diversification explain such differences.

Finally, the third chapter presents a two-country model where the government bonds issued by one country can be used to ease financial transactions globally, resulting in endogenous convenience yields for these assets. I find that the new issuance of convenience assets spills over to foreign households, as their equilibrium transaction costs are reduced. Moreover, a global liquidity shock affects both countries differently, as the pricing of convenience assets increases in this shock and allows the issuing country to reduce taxes. Finally, I study the asset pricing implications of convenience yields in light of existing puzzles.

The dissertation of Mariano Joaquin Palleja is approved.

Andrew Granger Atkeson

Matthew Saki Bigio Luks

Lee Ohanian

Pierre-Olivier Weill, Committee Chair

University of California, Los Angeles

2024

To Camila, the love of my life.

To my family, the most loving and supportive people ever.

To my friends, thanks for all the laughter that makes life so much better.

To my teachers, thanks for paving the path.

Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

I am most grateful to Pierre-Olivier Weill for his unconditional support and advice, and for being an outstanding human being. I thank the members of my committee, Saki Bigio, Andy Atkeson, and Lee Ohanian, for their encouragement and continuous feedback. This dissertation would not have been possible without their guidance. I also thank Mahyar Kargar, David Baqaee, and Joao Guerreiro, for their useful suggestions. These essays have also benefited from long discussions with Fatih Ozturk and Luis Cabezas, to whom I owe a huge amount of gratitude. I thank all the friends I met in graduate school. Their companionship made me feel at home.

The first and second chapters have also been enriched by extensive conversations with Yesol Huh, Michael Gordy, Jessica Li, Valery Polkovnichenko, Andreas Rapp, and Guillaume Rocheteau. I am deeply thankful for their generosity.

I would also like to thank Javier Garcia-Cicco, my first advisor back in Argentina, who supported and guided me during the first steps of this journey. The third chapter of this dissertation heavily relies on the tools he introduced to me.

VITA

Mariano Joaquin Palleja

Education

University of California, Los Angeles

M.A. in Economics 2018-2019

Universidad de San Andres

M.S. in Economics 2014-2015

Universidad de Buenos Aires

B.A. in Economics 2007-2012

Research in Progress

"Over-the-Counter Intermediation, Customers' Choice and Liquidity Measurement" (*Job Market Paper*)

"Portfolio Trading in OTC Markets: Transaction Cost Discounts and Penalties"

"An International Macro Model with Convenience Yields"

Publications

"Conditional Exchange Rate Pass-through: A DSGE Approach". Center for Latin American Monetary Studies, 2019.

Relevant Positions

University of California, Los Angeles

Teaching Fellow 2019-2023

Federal Reserve Board

Summer Dissertation Fellow – Risk Analysis, R&S Division 2023

Central Bank of Argentina

Economics Research Analyst - Research Department 2015-2018

Inter-American Development Bank

Winter Intern - Education Division 2015

Fellowships, Honors, and Awards

Best Proseminar in Monetary Economics and Macroeconomics, UCLA. 2023

Dissertation Year Fellowship – Graduate Division, UCLA. 2023-2024

Summer Dissertation Fellowship – Federal Reserve Board. 2023

Summer Fellowship – Department of Economics, UCLA. 2020

Graduate Division Fellowship – Department of Economics, UCLA. 2018-2023

Rodrigo Gómez Award, Center for Latin American Monetary Studies. 2018

Research funds and grants

Lewis L. Clarke Fellowship Fund, UCLA. 2023

Doctoral Student Travel Grant, UCLA. 2023

Presentations

Cornerstone, Universidad Catolica de Chile, Carlos III Business School, 2024
CUNEF, Commodity Futures Trading Commission, Office of the Comptroller of the Currency, U.S. Treasury Office of Financial Research, Central Bank of Chile, Joint Committee of Taxation, ORT Uruguay, Rice-LEMMA Monetary Conference.

Federal Reserve Board R&S Workshop, Econometric Society European Meeting, 2023
Northern Finance Association, West Coast Search and Matching Workshop.

CVII Meeting of Central Bank Governors - CEMLA. 2019

Central Bank of Argentina. 2018

Chapter 1

Over-the-Counter Intermediation, Customers' Choice and Liquidity Measurement

Stringent financial regulations and advancing trading technologies have reshaped over-the-counter intermediation, discouraging dealers from providing immediacy to customers using their own inventories (principal trades) in favor of a larger matchmaking activity (agency trades). This paper studies how customers optimally choose between these two trading mechanisms and the implications of this choice for market liquidity. I develop a quantitative search model where heterogeneous customers choose between immediate but expensive and delayed but less costly trades, i.e., principal and agency trades, respectively. Each customer solves this speed-cost trade-off, jointly determining her optimal mechanism, transaction costs, and trading volume. When market conditions change, customers migrate across mechanisms in pursuit of higher trading surpluses. I show that this migration is not random, thus liquidity measures change not only because of changes in market conditions but also because of a composition effect. To quantify such an effect, I structurally estimate my model and build counterfactual measures that control for migration. I replicate the major innovations seen in these markets and find that composition effects explain more than a third of the increase in principal transaction costs.

1.1 Introduction

Over-the-counter (OTC) markets are characterized by the lack of a centralized exchange in which customers can trade securities. Instead, customers need to search for trading counterparties. Dealers mitigate these search frictions in two ways. First, by trading with customers using their own inventories, i.e., by performing principal trades. Second, by matching customers with offsetting liquidity needs, i.e., by performing agency trades.[1] These two trading mechanisms, principal and agency, represent for customers a speed-cost trade-off. Principal trades are immediate but, given the implied inventory costs, are also costly. In contrast, agency trades are cheaper but imply an execution delay, caused by the time it takes to find a suitable counterparty.

Post-2008 financial regulations and recent technological changes have had a major impact on the relative cost of supplying these two types of trades. The implementation of the Dodd-Frank Act and the Basel III framework increased dealers' inventory costs, reducing their willingness to trade on a principal basis (Duffie, 2012; Bessembinder, Jacobsen, Maxwell, and Venkataraman, 2018). Quoting Goldman Sachs: *"Banks are committing less capital to trading desks with fixed income assets down 22% since 2010, and have exited some businesses altogether; for example, J.P. Morgan and Morgan Stanley no longer make mar-*

[1] Agency trades are also known in the literature as riskless principal or matchmaking trades. The key characteristic of this mechanism is that the dealer avoids involving her own inventories by pre-arranging both legs before executing them.

kets in physical commodities while Deutsche Bank has exited single-name CDS"[2]. In turn, the rising popularity of electronic trading venues made matching customers easier, shifting intermediation further away from dealers' inventories (O'Hara and Zhou, 2021).

Although the literature has extensively studied dealers' optimal intermediation strategy in the face of changing market conditions, the customers' optimal response to such a strategy and its implications for liquidity measurement have remained relatively unexplored. Notably, the speed-cost trade-off previously described suggests that customers may optimally migrate across trading mechanisms when market conditions change. Moreover, the decentralized nature of OTC markets – in which each customer bargains her own terms of trade – suggests that this migration might affect liquidity measures, by altering the samples over which these measures are computed.

In this paper, I develop and estimate a quantitative search model where I explicitly study the trading mechanism choice of each customer. I use this model to address how this trading mechanism choice affects liquidity measures when market conditions change. The model features risk-averse customers choosing between immediate but expensive and delayed but less costly trades, i.e., principal and agency trades, respectively. I find that customers with larger trading needs choose to buy and sell on principal. Intuitively, when trading is relatively urgent, the immediacy benefit outweighs the principal premium paid. Furthermore, customers with larger trading needs pay higher transaction costs, given that dealers extract higher fees from them. When market conditions change a fraction of customers optimally migrate across trading mechanisms. Therefore, principal and agency transaction cost measures change not only because the market conditions did, but also because of a composition effect. To quantify this composition effect, I develop counterfactual measures of transaction costs that control for migration. I structurally estimate the model using corporate bond transaction data and revisit the two major innovations this market experienced in the last

[2] Goldman Sachs Global Investment Research, August 2, 2015 Report.

decade. I find that the standard practice of comparing average transaction costs before and after a change in market conditions overestimates the impact of these changes. Specifically, composition effects account for 32% of the rise in principal costs after an inventory costs increase and for around 90% of the change after an increase in the agency execution speed. In turn, agency costs are barely affected by composition effects.

My model explicitly accounts for the optimal decisions of customers facing alternative trading mechanisms in OTC markets. Particularly, I build on the framework in Lagos and Rocheteau (2009) (hereafter LR09). The model features search frictions, heterogeneous risk-averse customers trading a perfectly divisible asset, and bilateral bargaining over the terms of trade. My theoretical contribution relative to LR09 is that I allow customers to choose between two trading mechanisms, which resemble principal and agency trades in practice. Principal trading is immediate but costly. This responds to dealers partially translating their implied inventory costs to customers. Agency trading is delayed but cheaper: finding a suitable counterparty takes time, but dealers avoid incurring inventory costs. These features enable me to study the aforementioned speed-cost trade-off.

I find that, in equilibrium, customers sort themselves across mechanisms depending on their liquidity needs. Customers with a larger distance between current and optimal asset positions choose to trade on principal. Conversely, customers with positions closer to their optimal ones choose to wait for an agency execution. This finding is explained by customers obtaining a marginally decreasing utility from holding assets. The bigger the distance between customers' current and optimal positions, the higher their marginal trading surplus and the higher their willingness to pay for an immediate execution.

This optimal sorting has a direct impact on liquidity measures. In the model, optimal mechanisms and transaction costs are jointly determined. Specifically, transaction costs are bargained, and thus they incorporate a customer's specific trading surplus. The more a customer needs to trade, the larger the marginal trading surplus she attains and the higher the cost she has to pay for each unit traded. As can be seen, when trading needs are large, not only are customers more likely to opt for the principal trade, but they also pay higher transaction costs. The implication is that principal traders pay on average higher costs not only because of the inventory costs implied by the mechanism but also because of selection: customers trading on principal have on average larger trading needs than those trading on agency.

I use this framework to analyze the optimal reaction of customers when market conditions change and its implications for liquidity measurement. Specifically, I consider changes in the two key parameters that affect the speed-cost trade-off faced by customers: the inventory costs implied by principal trades and the execution speed of agency trades. These changes resemble recent market innovations, where stricter regulations increased inventory costs and the rising popularity of electronic trading venues eased agency trading. Not surprisingly, in both cases, customers endogenously migrate away from principal trading. Furthermore, such migration is not random: among principal traders, only those with smaller trading needs migrate towards agency. Intuitively, smaller trading needs place customers closer to being indifferent between principal and agency trading, given that the marginal surplus from fast trading is closer to the premium cost paid for it.

Such a heterogeneous response implies an empirical issue when trying to estimate the impact of a market innovation on liquidity. In this regard, the empirical literature has widely exploited the relation between trading mechanisms and execution delays to overcome a recurrent inconvenience: execution delays are not observed. Particularly, when measuring transaction costs, researchers would split trades beforehand according to the trading mechanism used. Principal costs would account for the price of immediacy, whereas agency costs would measure the price of delayed executions [3]. Although splitting trades in such a way

[3] There are two main strategies to identify principal and agency trades. The first one infers agency trades

purges transaction cost measures from execution delay changes, it overlooks the fact that the obtained samples are endogenous: they are the result of a choice. When market conditions change, customers endogenously migrate, and thus the estimates of the impact on a mechanism's transaction costs are subject to a composition bias. For example, an increase in inventory costs would reduce the sample of principal traders to those with higher trading needs. In such a case, the effect of increasing inventory costs on principal transaction costs would be overestimated. This bias can hardly be narrowed when the characteristics in which the samples differ cannot be observed.

Equipped with the steady-state equilibrium of my model, I tackle this empirical issue. Firstly, I decompose the equilibrium distribution of customers into those that, after a market innovation, continue using the same mechanism or not, i.e., the non-migrant and migrant customers, respectively. Secondly, for each mechanism I compute measures of transaction cost changes, using both the entire distribution of customers before and after the innovation, as well as the subset of non-migrant customers. The comparison of these measures returns the sign and size of the composition bias.

To ensure that my numerical results are grounded in the data, I structurally estimate the model. For this, I use U.S. corporate bond secondary market transaction data. Specifically, I employ the academic version of the Trade Reporting and Compliance Engine (TRACE) database from January 2016 to December 2019. Importantly, this data contains dealers' identifiers, thus it allows me to distinguish between principal and agency trades. I target a set of relevant empirical moments and use the generalized method of moments to jointly estimate the deep parameters of the model.

as those offsetting transactions performed by the same dealer within a small time window (usually between one and fifteen minutes), labeling as principal all remaining trades (Schultz, 2017; Goldstein and Hotchkiss, 2020; O'Hara and Zhou, 2021; Choi, Huh, and Seunghun Shin, 2024). A second method is to isolate episodes where arguably only principal trades are performed, such as downgrades (Bao, O'Hara, and Zhou, 2018), extreme market volatility events (Anderson and Stulz, 2017), or index exclusions (Dick-Nielsen and Rossi, 2019).

Finally, the estimated model is used to revisit the empirical evidence related to the transaction costs evolution after two major OTC markets' innovations. I perform numerical exercises that replicate both the introduction of post-2008 stricter financial regulations and the rise of electronic trading venues. In both cases, when the economic environment changes, migration across mechanisms takes place. Using the aforementioned strategy, I show that the composition bias matters: it explains an economically significant fraction of the change in transaction costs.

Regarding the first exercise proposed, the aftermath of the 2008 financial crisis saw the introduction of new regulations aimed at increasing the financial market's resilience. The adoption of the Dodd-Frank Act in the United States and the Basel III framework internationally – regulations meant to reduce banks' exposure to risky assets – negatively affected their dealership activity. Specifically, these regulations increased banks' cost of holding assets in their balance sheets, thus reducing their willingness to provide liquidity on a principal basis (Duffie, 2012). Several papers have addressed the impact of these new regulations on market transaction costs. Overall, the consensus is that principal costs have increased since the new regulations took place, with intermediation shifting away from principal trading towards larger agency activity (Anderson and Stulz, 2017; Schultz, 2017; Bao, O'Hara, and Zhou, 2018; Bessembinder, Jacobsen, Maxwell, and Venkataraman, 2018; Dick-Nielsen and Rossi, 2019; Choi, Huh, and Seunghun Shin, 2024). I analyze such an increase in inventory costs through the lens of the model. The exercise suggests that previous estimates overstate the increase in principal costs. Particularly, I find that the composition bias accounts for a third of the increase in principal costs while it does not play an economically significant role in the change of agency costs.

The second numerical exercise is motivated by the emergence of electronic trading venues. Compared to traditional voice trading, electronic requests for quotes allow customers to contact multiple dealers simultaneously. The empirical evidence tells us that the agency share is higher for bonds that are traded electronically and that dealers use electronic platforms to find counterparties for customers that contacted them through traditional voice messages (Bessembinder, Jacobsen, Maxwell, and Venkataraman, 2018; O'Hara and Zhou, 2021). From the customers' perspective, the rising popularity of electronic trading venues implies that dealers can match them with a counterparty faster. To replicate this market innovation, I reduce the expected agency execution delay of the model. I find that transaction costs increase in both mechanisms. However, while the composition bias implies a negligible underestimation of the change in agency costs, it explains most of the increase in principal transaction costs.

Overall, the results in this paper suggest that accounting for customers' optimal response better informs policymakers about the impact that market innovations have on market liquidity. Firstly, this is because customers optimally migrate across mechanisms, mitigating the effect of worsening conditions and fostering the effects of improving ones. Secondly, considering the customers' response allows us to better measure the impact of the new market conditions. In particular, I show that stricter financial regulations have not increased principal transaction costs as much as was previously thought.

1.1.1 Related Literature

This paper develops a theoretical model of trading mechanism choice in OTC markets that allows me to revisit quantitatively recent evidence on transaction cost changes. It contributes to three strands of the literature.

Firstly, this paper contributes to the search literature in OTC markets, pioneered by Duffie, Gârleanu, and Pedersen (2005) and Lagos and Rocheteau (2009), and summarized in Weill (2020). In this literature, when customers and dealers meet, execution is immediate.

I relax this assumption by explicitly modeling two trading mechanisms, which resemble principal and agency trades in practice. This feature allows me to study theoretically the customers' trade-off between expensive but immediate and cheaper but slower execution. I show that the optimal mechanism choice can be characterized by preference-specific asset holdings thresholds, and analyze how such thresholds change according to the key parameters of the model. In their independent, contemporaneous work, Dyskant, Silva, and Sultanum (2023) also include alternative trading mechanisms in a search model. In their framework, customers are restricted to holding either zero or one unit of the asset. In contrast, I allow for unrestricted asset holdings and show that the endogenous trade size of each customer determines her trading mechanism choice. I further exploit the relation between trade size and transaction costs to estimate my model and perform quantitative exercises where I assess the role that migration plays when measuring liquidity.

This paper also contributes to the theoretical literature that explicitly accounts for principal and agency trading in OTC markets (Cimon and Garriott, 2019; Plante, 2021; An, 2022; An and Zheng, 2023; Saar, Sun, Yang, and Zhu, 2023). This literature addresses how dealers manage their inventories by setting the optimal principal trade cost: if the principal cost increases customers migrate towards agency trading, reducing the inventory burden [4]. In my model, both the trading mechanism choice and the terms of trade in each mechanism are the results of bilateral bargaining between dealers and customers. The consequences are twofold. First, it provides a non-degenerate distribution of transaction costs within each trading mechanism, which I exploit to estimate the model. This is because the terms of trade reflect both the incurred cost of the bargaining dealer and the trading surplus of the bargaining customer. Second, it allows me to study how composition effects affect liquidity measures in a quantitative way. In line with the existing literature, when

[4] A less related literature studies the customers' optimal choice of trading in a centralized or a decentralized market (Miao, 2006; Shen, 2015)

the principal premium increases the sample of customers trading on principal reduces. In contrast with the existing literature, the reduction of the sample does affect the average principal transaction costs, given that each customer bargains her own transaction cost.

Finally, this paper complements the empirical literature that addresses transaction cost changes and trading mechanism shifts in OTC markets. It has been documented that the regulation set after the 2008 financial crisis changed the liquidity profile of the corporate bond market. Specifically, researchers have shown that principal trading is less abundant and more costly (Anderson and Stulz, 2017; Schultz, 2017; Bao, O'Hara, and Zhou, 2018; Bessembinder, Jacobsen, Maxwell, and Venkataraman, 2018; Dick-Nielsen and Rossi, 2019; Choi, Huh, and Seunghun Shin, 2024; Rapp and Waibel, 2023). Additionally, the empirical evidence indicates that the rising popularity of electronic trading venues had attracted volume towards agency trading, reducing the cost of such trades (Bessembinder, Jacobsen, Maxwell, and Venkataraman, 2018; O'Hara and Zhou, 2021). Finally, during episodes of big turmoil, e.g., COVID-19, researchers have documented a rise in the cost of principal trading with an associated shift away from it (Kargar, Lester, Lindsay, Liu, Weill, and Zúñiga, 2021). A common feature across these papers is the lack of customer data, which prevents them from controlling the documented customers' endogenous migration when computing transaction cost changes [5]. I complement these papers by analyzing the sign and size of the consequent composition bias. To achieve this goal, I exploit the model to construct counterfactual distributions in which transaction cost changes can be measured using a steady sample of customers. I show that the estimates of transaction cost changes provided by this literature include an economically significant composition bias, and thus can hide the true speed-cost trade-off customers face.

[5] Goldstein and Hotchkiss (2020) address the cross-section of bond characteristics as another source of endogeneity. The authors find that bonds with an expected larger holding period are more likely to be traded on an agency basis, reconciling the fact that low turnover assets are often traded at smaller transaction costs.

1.2 The Model

In this section I explain the model. I start by describing the environment and the problems that both customers and dealers face. Later I show how terms or trade are set, highlighting the link between transaction costs and trading mechanism choice. Finally, I define the steady-state equilibrium.

1.2.1 Environment

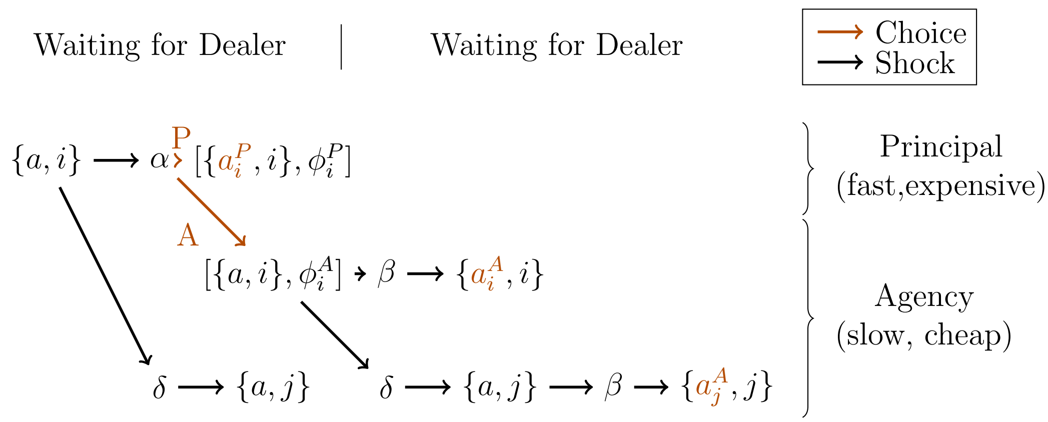
I build on LR09 continuous time model of an OTC secondary market with search frictions. There is a single asset in fixed supply $A \in \mathbb{R}_+$, and two types of infinitely lived agents: customers and dealers, both in unit measure and discounting time at rate $r > 0$. Customers hold an asset in quantity $a \in \mathbb{R}_+$ and derive utility from two different consumption goods, *fruit* and *numéraire*. *Fruit* is perishable, non-tradable, and produced by the asset in a one-to-one ratio. In turn, the *numéraire* good is produced by all agents. The instantaneous utility function of a customer is $u_i(a) + d$, where a and d represent the consumption of *fruit* and the net consumption of the *numéraire* good, respectively, and $i \in \{1, ..., I\}$ indexes the preference type. Specifically, the instantaneous utility provided by *fruit* is assumed iso-elastic, $u_i(a) = \epsilon_i \times a^{1-\sigma}/(1-\sigma)$, with multiplicative preference shifters ϵ_i. Each customer is subject to an independent preference shock process, which follows a Poisson distribution with arrival rate δ. Once hit by the preference shock, a new type i is assigned with probability π_i, where $\sum_{i=1}^{I} \pi_i = 1$. This change in preferences creates a motive for trade in the model, and can be interpreted as changing hedging needs (Duffie, Gârleanu, and Pedersen, 2007; Vayanos and Weill, 2008), changing beliefs about the asset's future payoff (Hugonnier, 2012), etc.

Customers can trade assets only when they contact a dealer, an event that is governed by a Poisson process with an arrival rate of α. Once a customer meets a dealer, she chooses among two kinds of trading mechanisms: principal or agency, denoted by superscripts P and A, respectively. On the one hand, if she opts for the principal trade, she immediately exchanges each unit of her excess position at the inter-dealer price p and pays a transaction cost of ϕ^P. On the other hand, if she opts for an agency trade, she waits until the dealer finds her a counterparty, and meanwhile enjoys the utility provided by her current asset holdings. It is assumed that she will be matched at a random time according to a Poisson process with β arrival rate. When matched, this customer rebalances her position at p and pays the dealer a transaction cost ϕ^A. I further assume that a customer cannot contact any other dealer while she is waiting for her trade to be executed. Thus, at every moment, customers will be either waiting to contact a dealer or waiting for their agency trade to be executed. These two states are denoted by ω_1 and ω_2, respectively.

Transaction costs and quantities are determined through a Nash bargaining protocol that takes place at the moment of contact with the dealer. This timing assumption implies that, for agency trades, the negotiation is based on the expected trade surplus a customer subject to preferences shocks might achieve. More details about these terms of trade are presented in subsection 1.2.2. After transactions are completed, the dealer and the customer part ways.

At any time, customers find themselves with certain asset holdings a_t, preference type i_t, and within a specific waiting state ω_t. Thus, customers can be fully characterized by the triplet $\{a_t, i_t, \omega_t\} \in \mathcal{O}$, where $\mathcal{O} = \mathbb{R}_+ \times \{1, ..., I\} \times \{\omega_1, \omega_2\}$. This heterogeneity is depicted with a probability space $(\mathcal{O}, \Sigma, H_t)$, where Σ is the σ-field generated by the sets $(\mathcal{A}, \mathcal{I}, \mathcal{W})$, with $\mathcal{A} \subseteq \mathbb{R}_+$, $\mathcal{I} \subseteq \{1, ..., I\}$, $\mathcal{W} \subseteq \{\omega_1, \omega_2\}$, and H_t is a probability measure on Σ that represents the distribution of customers across the state space at time t. Figure 1.1 outlines a customer's potential paths from the moment she contacts a dealer until she executes her trade.

Figure 1.1: Customer Path.



Note: This figure shows a customer's path through the state space. Shocks are depicted by black arrows, and include the contact with dealers (α), the change of preference (δ), and the execution of the agency trade (β). The customer's choice is depicted in orange arrows and includes the optimal trading mechanism and the corresponding new asset holdings.

Since I am going to focus on the steady-state equilibrium, to simplify the notation I disregard the time dependence when it is not strictly necessary. The maximum expected discounted utility attainable by a customer waiting for a dealer with preference type i at time t and asset holding a, $V_{i(t)}(a)$, satisfies

$$V_{i(t)}(a) = \mathbb{E}_{i(t)}\Big[\int_t^{T_\alpha} e^{-r(s-t)} u_{i(s)}(a)ds + e^{-r(T_\alpha - t)} \max\Big\{V^P_{i(T_\alpha)}(a), V^A_{i(T_\alpha)}(a)\Big\}\Big], \tag{1.1}$$

where

$$V^P_{i(T_\alpha)}(a) = V_{i(T_\alpha)}(a^P_{i(T_\alpha)}) - p(a^P_{i(T_\alpha)} - a) - \phi^P_{i(T_\alpha)}(a),$$
$$V^A_{i(T_\alpha)}(a) = \int_{T_\alpha}^{T_\beta} e^{-r(s-T_\alpha)} u_{i(s)}(a)ds + e^{-r(T_\beta - T_\alpha)}\Big[V_{i(T_\beta)}(a^A_{i(T_\beta)}) - p(a^A_{i(T_\beta)} - a) - \phi^A_{i(T_\alpha)}(a)\Big].$$

T_α and T_β are the next time a customer contacts a dealer and the execution time of the agency trade, respectively. The expectation operator $\mathbb{E}_{i(t)}$ is over the arrival times of contact with dealers, the execution of the agency trade, and the expected stream of preference types

$i(s)$, conditional on the customer being of a certain preference type at t. Transaction costs and prices are expressed in units of the *numéraire* good.

Note that the optimal asset holdings under the two trading mechanisms, $a^P_{i(T_\alpha)}$ and $a^A_{i(T_\beta)}$, might differ for two reasons. Firstly, a customer might change her type during the waiting period of a delayed trade. Hence, types $i(T_\alpha)$ and $i(T_\beta)$ might be different. Secondly, the transaction costs charged by dealers in each kind of trade might differ independently of the aforementioned reason: each trading mechanism will require the dealer to face a different cost. Since transaction costs add up to the effective price of a trade, customers may choose different optimal asset holdings in different mechanisms.

In turn, dealers trade on behalf of their customers in the inter-dealer market. If they are asked to execute a principal trade, they need to incur a cost $\theta \in [0, \frac{r}{r+\beta})$ per (*numeraire*) dollar traded. In line with existing literature (e.g., An and Zheng, 2023; Saar, Sun, Yang, and Zhu, 2023), I assume that dealers' marginal inventory costs are constant. In this regard, Duffie et al. (2023) shows that liquidity measures are not affected by the level of dealers' inventory capacity utilization unless the latter is at an abnormally high level. Thus, the assumption is empirically supported as such a scenario of extremely high capacity utilization is not considered [6]. On the other hand, if the client asks the dealer to perform an agency trade, they wait until a counterparty is found, and the transaction cost is charged at execution. A representative dealer does not hold positions and her instantaneous utility equals her consumption of the *numéraire* good. Thus, her expected utility is given by the present value of the transaction costs she collects net of the costs she incurs. A dealer's maximum

[6] In terms of modeling choice, this reduced form formulation allows a link to be drawn between the demand for immediacy and dealers' inventory costs without dealing with inventories as an additional state variable. See Cohen, Kargar, Lester, and Weill (2022) for a search model with explicit inventory in OTC markets.

expected discounted utility satisfies

$$W(t) = \mathbb{E}\Big[e^{-r[T_\alpha - t]}\Big(\int_{\mathcal{O}} \Phi_{i(T_\alpha)}(a) dH_{T_\alpha} + W(T_\alpha)\Big)\Big], \tag{1.2}$$

where $\Phi_i(a) = \mathbf{1}_{[\text{P trade}]}\Big(\phi_i^P(a) - \theta p|a_i^P - a|\Big) + \mathbf{1}_{[\text{A trade}]}\Big(e^{-r(T_\beta - T_\alpha)}\phi_i^A(a)\Big)$ and the integration over the probability measure H_{T_α} is because of random matching.

1.2.2 Terms of Trade

In the proceeding subsections I derive the policy functions of the agents of the model, i.e., the optimal asset holdings, their corresponding transaction costs, and the trading mechanism choices. I find that, in equilibrium, customers sort across mechanisms depending on their liquidity needs.

Optimal Asset Holdings and Transaction Costs

Once a customer contacts a dealer and chooses a trading mechanism, optimal asset holdings and transaction costs are set as the outcome of a Nash bargaining problem, where the dealer's bargaining powers is $\eta \in [0, 1]$ [7]. When trading on principal, execution is immediate, and so the trade surplus of the customer equals the utility gains of re-balancing positions minus the total price paid for it. On the dealer's side, her trade surplus equals the transaction cost charged minus the cost of performing principal trades. Hence, the Nash product writes

$$\{a_i^P(a), \phi_i^P(a)\} = \arg\max_{(a', \phi')} \Big\{V_i(a') - V_i(a) - p(a' - a) - \phi'\Big\}^{1-\eta}\Big\{\phi' - \theta p|a' - a|\Big\}^{\eta}.$$

[7] Duffie, Gârleanu, and Pedersen (2007) model explicitly a bargaining game where agents make alternate offers. They show that the Nash bargaining powers equal the probabilities of making an offer in such a game.

The solution for optimal principal terms of trade is

$$\phi_i^P(a) = \eta\big[V_i(a_i^P(a) - V_i(a) - p(a_i^P(a) - a)\big] + (1-\eta)\big[\theta p|a_i^P(a) - a|\big], \tag{1.3}$$

$$a_i^P(a) = \arg\max_{a'} \quad V_i(a') - V_i(a) - p(a' - a) - \theta p|a' - a|. \tag{1.4}$$

The presence of inventory costs has two important consequences for principal trades. Firstly, conditional on the trade direction, inventory costs are translated into an increase (decrease) in the effective price customers pay when buying (obtain when selling). Thus, the problem becomes linear in the volume traded, and consequently, customers choose their optimal holdings independently of their current positions. Secondly, some customers might optimally not trade at all. In contrast with LR09 and the bulk of theoretical models that account for principal and agency trades, the policy function in the model allows for a no-trade region, explained by the existence of immediacy costs[8]. Whenever the gain in lifetime utility minus the inter-dealer price paid for such trade does not outweigh the immediacy costs, it is better not to trade on a principal basis. Furthermore, if keeping the current position is preferred over engaging in an agency trade, the optimal policy is not to trade at all.

These two consequences can be easily seen by optimizing Equation (1.4) conditional on the trade direction a principal trader would pursue. Particularly, current asset holdings can be partitioned into three subsets, which I denote by $\Gamma_i \in \{Buy_i, Sell_i, NoT_i\}$:

[8]Given that most of the databases are based on transaction data, the empirical evidence related to no trades is hard to find. Hendershott, Li, Livdan, and Schürhoff (2020) provide evidence of no trading in the CLO market. The authors compute a no-trading rate that goes from 7% to 30%, decreasing in the seniority tranche of the security. The CLO market features, in which trading is done through auctions and where sellers choose when to contact dealers, prevent us from reading these numbers through the lens of the present model.

$$
\Gamma_i = \begin{cases} Buy_i: & a \,|\, [V_i(a') - a'p] - [V_i(a) - ap] > \theta p(a' - a) \quad \text{for some } a' \in (a, \infty), \\ Sell_i: & a \,|\, [V_i(a') - a'p] - [V_i(a) - ap] > \theta p(a - a') \quad \text{for some } a' \in [0, a), \\ NoT_i: & a \,|\, [V_i(a') - a'p] - [V_i(a) - ap] \leq \theta p|a' - a| \quad \forall a' \neq a. \end{cases}
$$

Within each subset, optimal asset holdings can be easily characterized[9]:

$$
a_i^P(a) = \begin{cases} a_i^{P,b} = \arg\max_{a'}\{V_i(a') - p(1+\theta)a'\} & \text{if } a \in Buy_i, \\ a_i^{P,s} = \arg\max_{a'}\{V_i(a') - p(1-\theta)a'\} & \text{if } a \in Sell_i, \\ a & \text{if } a \in NoT_i, \end{cases}
$$

In turn, agency trades imply an expected execution delay, during which the customer might suffer a preference shock. Hence, a specific timing assumption regarding when optimal holdings and transaction costs are set is needed. In this regard, it is assumed that transaction costs are arranged when customers and dealers meet, and that optimal holdings are decided at execution. The implications of this assumption are twofold. Firstly, the model allows for order cancellation, a common practice when trading securities (Foucault, Pagano, and Röell, 2013). Secondly, agency transaction costs are set based on the expected gains from trade of customers who may suffer preference shocks while waiting[10].

[9] If the value function is increasing and strictly concave in asset holdings, these subsets are convex and the maximizers are unique. I check numerically both the convexity of the sets as well as the uniqueness of the maximizers and they both hold robustly.

[10] An alternative modeling choice is to assume that customers and dealers commit upon contact to trade a certain optimal volume at execution. In this case, an amplification of the effect presented in LR09 would be observed, where optimal asset holdings would be partially chosen according to the type at the moment of trading and partially according to their expected flow of types. If customers opt for agency trading, they choose their positions taking into account that they might change their preferences both before and after the execution of the trade, so the expected flow of types weight will be larger. This assumption not only is at odds with order cancellation in practice but also implies a modeling disadvantage. In particular, it requires tracking the committed trade amount within the "waiting for execution" state, adding another state variable to an already large state-space. Another alternative is to assume that the optimal volume traded and transaction costs are decided at execution. In that case, the utility that the agent loses from not having an optimal position during the waiting time would be a sunk cost and it would not be considered in the bargaining process or in the consequent terms of trade.

A customer's expected agency trade surplus is composed by two terms. The first component is her expected utility derived from holding her current position while waiting for execution. The second component is her expected future gains from re-balancing her position. On the dealers' side, their trade surplus is just the discounted transaction cost collected. Terms of trade when agency is chosen are set according to

$$
\begin{aligned}
&\{\{a_i^A\}_{i=1}^I, \phi_{i(t)}^A(a)\} \\
&= \underset{\{a_i''\}_{i=1}^I, \phi''}{\arg\max} \Big\{\mathbb{E}_{i(t)}\Big[\int_t^{T_\beta} e^{-r(s-t)} u_{i(s)}(a)ds + e^{-r(T_\beta - t)}\big[V_{i(T_\beta)}(a''_{i(T_\beta)}) - p(a''_{i(T_\beta)} - a) - \phi''\big]\Big] \\
&\qquad\qquad\qquad\qquad\qquad\qquad - V_{i(t)}(a)\Big\}^{1-\eta}\Big\{\mathbb{E}_t\big[e^{-r(T_\beta - t)}\phi''\big]\Big\}^{\eta}.
\end{aligned}
$$

The optimal terms in the agency trade are

$$
\begin{aligned}
\mathbb{E}_t[e^{-r(T_\beta - t)}]\phi_{i(t)}^A(a) = \eta\Big\{\mathbb{E}_{i(t)}\Big[\int_t^{T_\beta} e^{-r(s-t)} u_{i(s)}(a)ds \\
+ e^{-r[T_\beta - t]}\big[V_{i(T_\beta)}(a^A_{i(T_\beta)}) - p(a^A_{i(T_\beta)} - a)\big]\Big] - V_{i(t)}(a)\Big\}, \quad (1.5)
\end{aligned}
$$

$$
a_i^A = \underset{a''}{\arg\max} \quad \{V_i(a'') - pa''\}. \quad (1.6)
$$

With these results at hand, I manipulate the Bellman equation (1.1) to reach a simpler and more intuitive representation. First, I plug in the bargaining outcomes and note that the problem is equivalent to the one faced by a customer with maximum bargaining power but smaller contact rate $\kappa = \alpha(1-\eta)$. I refer to κ as the bargaining-adjusted contact rate. Second, I use analytical expressions for all the expectations related to the shocks of the model.[11]

$$V_i(a) =$$

$$\bar{U}_i^{\kappa}(a) + \hat{\kappa}\big[[1-\hat{\delta}^{\kappa}]\max\big\{V_i(a_i^P) - p(a_i^P - a) - \theta p|a_i^P - a|\,,\, \bar{U}_i^{\beta}(a) + \hat{\beta}[\bar{V}_i^A - p(\bar{a}_i^A - a)]\big\}$$
$$+ \hat{\delta}^{\kappa}\sum_j \pi_j \max\big\{V_j(a_j^P) - p(a_j^P - a) - \theta p|a_j^P - a|\,,\, \bar{U}_j^{\beta}(a) + \hat{\beta}[\bar{V}_j^A - p(\bar{a}_j^A - a)]\big\}\big], \tag{1.7}$$

where

$$\bar{U}_i^{\nu}(a) = \Big[[1-\delta^{\nu}]u_i(a) + \delta^{\nu}\sum_j \pi_j u_j(a)\Big]\frac{1}{r+\nu}$$

$$\bar{V}_i^A = [1-\delta^{\beta}]V_i(a_i^A) + \delta^{\beta}\sum_j \pi_j V_j(a_j^A) \quad , \quad \bar{a}_i^A = [1-\delta^{\beta}]a_i^A + \delta^{\beta}\sum_j \pi_j a_j^A$$

$$\hat{\kappa} = \frac{\kappa}{r+\kappa} \quad , \quad \hat{\beta} = \frac{\beta}{r+\beta} \quad , \quad \hat{\delta}^{\nu} = \frac{\delta}{r+\delta+\nu} \quad , \quad \nu = \{\kappa, \beta\}.$$

The first term of Equation (1.7), $\bar{U}_i^{\kappa}(a)$, is the expected utility of holding assets a until the next (bargaining-adjusted) contact with a dealer. While waiting for this contact, a customer might change her preferences, and so this term is a convex combination of the utility under the current and the future expected type. Hence, when the customer contacts a dealer she might be in two different situations: she might have avoided the preference shock or she might have received it. The corresponding probabilities of these scenarios are $(1-\hat{\delta}^{\kappa})$ and $\hat{\delta}^{\kappa}$, respectively.

If customers choose to trade on principal, the execution is immediate. The premium paid for such immediacy is expressed in a higher effective price for buyers, $p(1+\theta)$, and a lower effective price for sellers, $p(1-\theta)$. Conversely, if an agency trade is chosen, customers need to wait for execution. This waiting stage is reflected in $\bar{U}_i^{\beta}(a)$, the utility that a customer

[11] See the Appendix 1.A.2 and 1.A.3 for a step-by-step computation.

with current preference i holding asset a expects to derive until executing her agency trade. At the moment of execution, her preference may have changed, and so her expected value function, $\bar{V}_i^A$, is a convex combination across the preference space.

Equation (1.7) highlights the two differences between trading mechanisms. The first one is the expected execution delay that agency trading implies. The second one is the less favorable trading terms that customers face under principal trading, given the partial translation of dealers' inventory costs. These two differences define the trade-off that customers will have to solve.

Trading Mechanism Choice

When customers contact dealers, they must choose between an immediate principal or a delayed agency trade. I start by looking for the preference-specific current asset holding thresholds that make each customer indifferent among trading mechanisms. The indifference condition for a type i customer is given by:

$$\left[V_i(a_i^P) - V_i(a)\right] - p(a_i^P - a) - \theta p|a_i^P - a| = \left[\bar{U}_i^{\beta}(a) + \hat{\beta}\bar{V}_i^A - V_i(a)\right] - \hat{\beta}p(\bar{a}_i^A - a), \quad (1.8)$$

This equation compares the trade surplus in each mechanism, which are functions of customers' difference between their current and their optimal asset holdings. To gain intuition, Figure 1.2 graphs, for a mid-preference customer, these trade surpluses.

Figure 1.2 presents two salient features. First, as current and optimal asset holdings get closer, the principal surplus goes to zero but the agency surplus remains at a positive level. This is explained because principal trading is immediate, whereas agency trading is delayed. When a customer holds the optimal principal position given her current preference, a_i^P, trading on principal would represent no surplus: the optimal position is already achieved. However, when a customer holds the optimal agency position according to her current preferences, a_i^A, trading on agency might still represent a positive expected surplus. This is because, while customers wait for execution, her preferences might change making her current position no longer optimal.

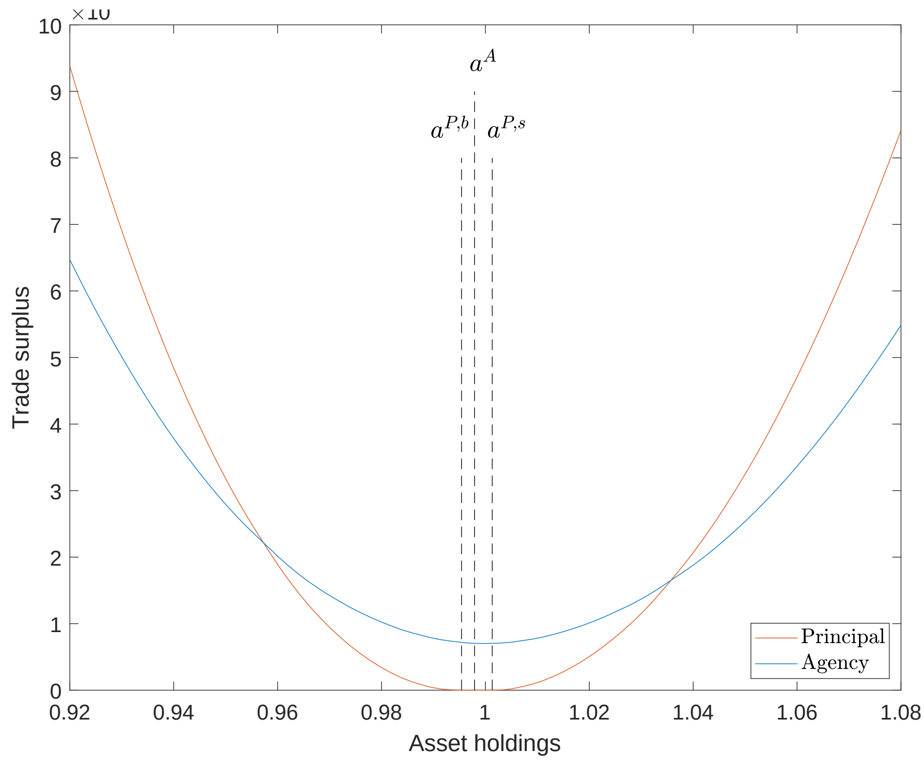
Second, customers with a larger distance between current and optimal asset holdings trade on principal. To analyze this pattern, let me consider a customer who compares whether to buy on principal or to engage in the agency trade. To further simplify the exposition, consider the limiting case where preference shocks arrive with a Poisson intensity close to zero, thus $\bar{U}_i^{\beta}(a) + \hat{\beta}\bar{V}_i^A = \frac{u_i(a)+\beta V_i(a_i^A)}{r+\beta}$ and $\bar{a}_i^A = a_i^A$. Equation (1.8) can be written:

$$\underbrace{\left[\frac{rV_i(a_i^A) - u_i(a)}{r+\beta}\right]}_{\text{cost of delay}} = \underbrace{p(1+\theta-\hat{\beta})(a_i^A - a)}_{\text{effective price diff}} + \underbrace{[V_i(a_i^A) - pa_i^A] - [V_i(a_i^P) - pa_i^P]}_{\text{gains from trade diff}} - \underbrace{p\theta(a_i^A - a_i^P)}_{\text{adjustment}}$$

The LHS expresses the cost of performing agency trades: while waiting for a suitable counterparty the customer will hold an unwanted position. The RHS expresses the benefits of performing agency trades. It is composed of three terms. First, agency trading allows avoiding inventory costs, and so the effective price paid is lower. Second, given that the effective price of trading on agency is more convenient than that of principal trading, a customer would trade a larger quantity in the former mechanism than in the latter. Finally, the transaction cost difference needs to be adjusted for the fact that, if the customer had traded on principal, she would have bought a smaller quantity, hence the total transaction cost difference paid to dealers would have been smaller.

The comparison between the costs and benefits of trading on agency tells us why customers with larger trading needs choose principal trades. Given a customer's preference type, only the first terms of both sides of the equation are affected by her current asset holdings. As the distance between current and optimal asset holdings increases, the cost

Figure 1.2: Trading mechanism choice.



Note: This figure depicts the trade surplus under the two trading mechanisms, for a customer with preference type at the center of the distribution. The optimal asset holdings under the principal trade, for buyers and sellers, are graphed in dashed lines. The values correspond to the baseline calibration presented in section 1.5.3

of delaying the execution increases at a faster rate than the savings given by the effective price difference. This is because the cost of each extra unit away from the optimal position is marginally increasing (utility is strictly concave), whereas the effective price difference is constant.[12]

I summarize the optimal trading mechanism rule for a customer with preference i and asset holdings a using the asset holding subset $\{\Gamma_i^P, \Gamma_i^A\}_{i=1}^I$. These are partitions of the subsets $\Gamma_i = \{Buy_i, Sell_i, NoT_i\}$, which in turn defined what the optimal trading direction was

[12]Note that, if preference shocks arrive at a positive rate, the logic follows: customers compare the costs of a delayed execution and the accumulated savings from the difference in effective prices, both terms only being affected by her current asset holdings.

for a customer trading on principal. This decision follows from the fact that the indifference equation (1.8) considers the optimal asset position in each mechanism and that the principal optimal position changes with the trade direction, as it was explained in subsection 1.2.2. In Appendix 1.A.4 I provide a discussion of how these sets are built.

1.2.3 Steady-state Distribution and Market Clearing

In this subsection I derive the general equilibrium steady-state equations of the model. As previously stated, a customer can be fully characterized by the triplet $\{a, i, \omega\}$. Thus, I first develop the equations needed to compute the steady-state distribution $H(a, i, \omega)$ over such individual states. Second, I state the market clearing condition to solve for the steady-state equilibrium price p.

Given that the model allows for the possibility of optimally not trading, potentially any initial asset holding $a \in R_+$ might be included in the ergodic set. In such a case, the steady-state equilibrium will be conditioned by the initial holdings of assets across customers. In order to prevent such a pathological case, I focus on calibrations where $\cap_{i=1}^{I} NoT_i^P = \emptyset$. In other words, I focus on equilibria where there is no asset position such that every type decides not to trade when holding it[13]. Under this restriction, given that $\pi_i > 0\,\forall i$, every customer with any asset holdings will eventually trade. Hence, in the steady state, a customer will hold assets $a \in \mathcal{A}^*$, where $\mathcal{A}^* = \cup_{i=1}^{I}\{a_i^{P,b}, a_i^{P,s}, a_i^{A}\}$, and the steady-state distribution is characterized by the vector $n_{[a,i,\omega]}$. Equations (1.4) and (1.6) provide the optimal asset position in each kind of trade, and subsets $\{\Gamma_i^P, \Gamma_i^A\}_{i=1}^{I}$, with $\Gamma = \{Buy, Sell, NoT\}$, indicate which kind of trade customers wish to perform. These policy functions and the three shocks present in the model indicate how to track customers across the discrete state space. Since,

[13] As will be explained in section 1.5, the GMM procedure used to estimate the model searches through the parametric space in an unrestricted manner, yielding a calibration where the restriction here imposed is not binding

in the steady state, the flow of customers entering and exiting each individual state should be equal, the following set of inflow-outflow equations computes the stationary distribution of the model.

$$n_{[a_i^{P,b},i,\omega_1]}: \quad \delta\pi_i \sum_{j\neq i} n_{[a_i^{P,b},j,\omega_1]} + \alpha \sum_{a\in Buy_i^P} n_{[a,i,\omega_1]} = n_{[a_i^{P,b},i,\omega_1]}\big(\delta(1-\pi_i) + \alpha \mathbf{1}_{[a_i^{P,b}\notin NoT_i^P]}\big) \tag{1.9}$$

$$n_{[a_i^{P,s},i,\omega_1]}: \quad \delta\pi_i \sum_{j\neq i} n_{[a_i^{P,s},j,\omega_1]} + \alpha \sum_{a\in Sell_i^P} n_{[a,i,\omega_1]} = n_{[a_i^{P,s},i,\omega_1]}\big(\delta(1-\pi_i) + \alpha \mathbf{1}_{[a_i^{P,s}\notin NoT_i^P]}\big) \tag{1.10}$$

$$n_{[a_i^{A},i,\omega_1]}: \quad \delta\pi_i \sum_{j\neq i} n_{[a_i^{A},j,\omega_1]} + \beta \sum_{a\in \mathcal{A}^*} n_{[a,i,\omega_2]} = n_{[a_i^{A},i,\omega_1]}\big(\delta(1-\pi_i) + \alpha \mathbf{1}_{[a_i^{A}\notin NoT_i^P]}\big) \tag{1.11}$$

$$n_{[a,i,\omega_1]}: \quad \delta\pi_i \sum_{j\neq i} n_{[a,j,\omega_1]} = n_{[a,i,\omega_1]}\big(\delta(1-\pi_i) + \alpha \mathbf{1}_{[a\notin NoT_i^P]}\big), \quad a \in \cup_{j\neq i}\{a_j^{P,b}, a_j^{P,s}, a_j^{A}\} \tag{1.12}$$

$$n_{[a,i,\omega_2]}: \quad \delta\pi_i \sum_{j\neq i} n_{[a,j,\omega_2]} + \alpha n_{[a,i,\omega_1]} \mathbf{1}_{[a\in\Gamma_i^A]} = n_{[a,i,\omega_2]}\big(\delta(1-\pi_i) + \beta\big), \quad a\in\mathcal{A}^* \tag{1.13}$$

The left-hand side of these equations represents the inflow in a specific individual state, and the right-hand side represents the outflow. As Figure 1.1 shows, in any time interval, three kinds of forces might move customers across states. Let us first consider the preference shock. The mass of customers of an individual state with preference i increases whenever customers from other states, with the same asset holdings and in the same waiting stage, receive the preference shock i. This happens with Poisson intensity $\delta\pi_i$. Similarly, that mass of customers decreases whenever customers therein are hit by preference shocks other than i. This happens with intensity $\delta(1-\pi_i)$. Second, let us consider the contact with dealer shock. This shock is received only by people waiting for a dealer, i.e., by customers within states where $\omega = \omega_1$, and happens with intensity α. Customers with current asset holdings that make them want to buy (sell) on principal will flow towards the state in which optimal asset holdings for principal buyers (sellers) correspond with their preference type. On the contrary, a customer with current asset holdings such that she opts for an agency trade will flow towards the waiting-for-execution stage, i.e., $\omega = \omega_2$, keeping both her holdings and preference type. It is worth noting that not all customers hit by this shock would travel across the state space. If a customer chooses not to trade, then she will remain in her current state until a preference shock eventually hits her. Finally, the execution shock, which happens with intensity β, moves customers across waiting stages. Obviously, such shock is received only by customers waiting for the execution of their trades, i.e., in states where $\omega = \omega_2$. Once a customer gets her agency trade executed, she goes back to the "waiting for dealers" stage. Since customers decide on optimal holdings at the moment of execution, this shock will move customers toward the state in which optimal agency asset holdings correspond with their preference type.

The set of Equations (1.9)-(1.13) can be represented by a transition matrix $T_{[3I\times I\times 2]}$, with attached transition probabilities $\pi^T_{n,n'}$, which denote the probability of moving from a state n towards a state n' in a given time length. Such a transition matrix can be used to update the vector of individual states masses until reaching the unique limit invariant distribution $n = \lim_{k\to\infty} n_0 T^k$, where n_0 is any initial distribution. Th.11.4 in Stokey, Lucas, and Prescott (1989) provides the conditions for this convergence result [14]. Once solved for the stationary distribution, the market clearing equation can be computed, and thus the steady-state equilibrium price p can be found. Aggregate gross demand in this secondary

[14] Basically, there should exist at least one state that receives inflows from all states with strictly positive probability. A sufficient condition for this to happen is that there exists a type i and a type j such that $\mathcal{A}^*_i \in Buy^P_j$, $\mathcal{A}^*_i \in Sell^P_j$ or $\mathcal{A}^*_i \in Buy^A_j \cap Sell^A_j$, where $\mathcal{A}^*_i = [a^{P,b}_i, a^{P,s}_i, a^A_i]$. Firstly, $\pi_i > 0\,\forall i$ and $\delta \in (0,1)$; therefore all types can turn into type i. Secondly, after customers of type i execute their trades, they go back to the waiting for a dealer stage. Finally, the condition described guarantees that, when those customers contact a dealer with their preferences i intact, they choose the same trading mechanism and eventually obtain the same optimal asset position. Thus such latter individual state would receive inflows directly or indirectly from all individual states. I check numerically and this condition robustly holds.

market is given by the weighted sum of individual states demands. Aggregate gross supply, in turn, is fixed by A. Therefore, the equilibrium price is the one at which the following market clearing equation holds:

$$\sum_{h=1}^{2} \sum_{i=1}^{I} \sum_{a \in \mathcal{A}^*} a n_{[a,i,\omega_h]} = A. \tag{1.14}$$

Note that, in the steady state, trading occurs constantly but the aggregate asset position is held constant. Given that all trades are cleared in the inter-dealer market, the market clearing condition (1.14) implies that the inter-dealer market is at equilibrium at all times. Of course, our steady state allows for a situation where the excess of demand in one mechanism is compensated by an excess of supply in the other.

1.2.4 Equilibrium

An equilibrium for this model is defined as a list of optimal asset holdings $\{a_i^P(a), a_i^A\}_{i=1}^I$, transaction costs $\{\phi_i^P(a), \phi_i^A(a)\}_{i=1}^I$, trading mechanism sets $\{\Gamma_i^P, \Gamma_i^A\}_{i=1}^I$ where $\Gamma = \{Buy, Sell, NoT\}$, stationary distribution $n_{[a,i,\omega]}$ and price p such that $\{a_i^P(a), a_i^A\}_{i=1}^I$ satisfies (1.4) and (1.6), $\{\phi_i^P(a), \phi_i^A(a)\}_{i=1}^I$ satisfies (1.3) and (1.5), $\{\Gamma_i^P, \Gamma_i^A\}_{i=1}^I$ are defined using thresholds satisfying (1.8), $n_{[a,i,\omega]}$ satisfies (1.9)-(1.13), and p satisfies (1.14).

In contrast with LR09, where the equilibrium can be found analytically, the model here presented needs to be solved numerically. The main difference with respect to LR09 in this regard is that current asset holdings affect not just the optimal portfolio, but also the trading mechanism chosen. To solve for the steady state of the model for any given inter-dealer price, p, I rely on the value function iteration method, enhanced with Howard's improvement step [15]. This procedure returns the policy and value functions conditional on p.

[15] See Appendix 1.A.5 for the necessary and sufficient conditions to use value function iteration as the solution method.

In turn, these functions are nested within the computation of Equation (1.14), which solves the inter-dealer price that clears the market in the steady state. The algorithm is described in detail in Appendix 1.A.6.

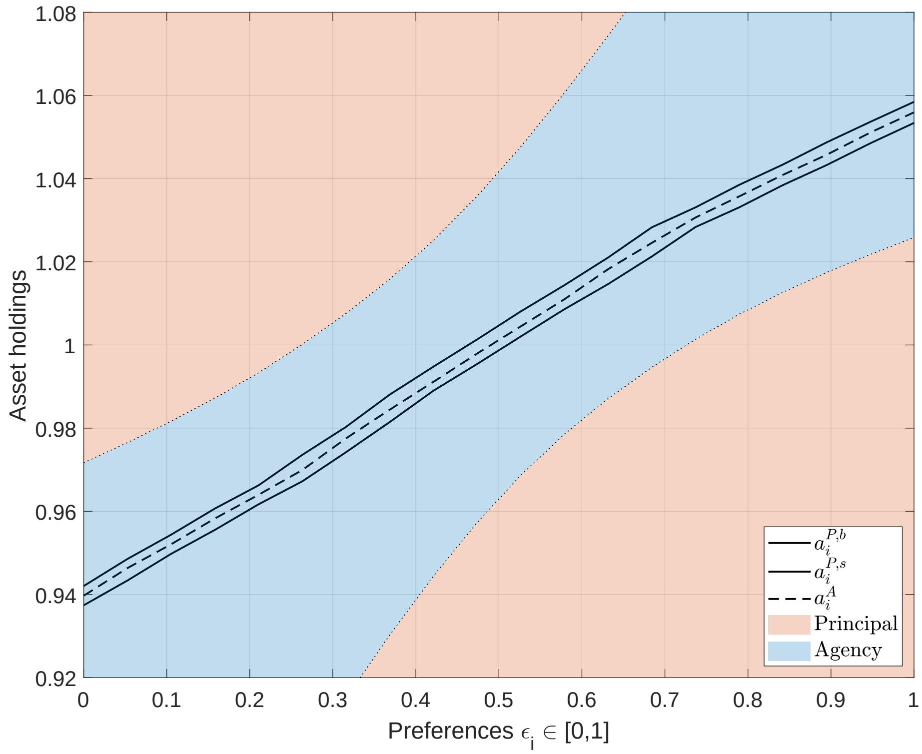
1.3 Equilibrium Allocations

In this section I study numerically the policy functions of the model. I use the parameter values that will be estimated in section 1.5. I initially map customers' preferences and current asset holdings with their optimal asset holdings and mechanism choices. I show that customers sort themselves across trading mechanisms according to their trading needs. After characterizing the pool of trades in each mechanism, I describe how such characteristics are translated into the transaction costs customers pay.

1.3.1 Equilibrium Asset Holdings and Trading Mechanism

The policy functions are presented in Figure 1.3. For each asset holding and preference type pair, $\{a, i\}$, I compute both the optimal asset holdings conditional on the trading mechanism and the trading mechanism choice. Regarding the optimal asset holdings, the lower and upper solid lines represent the buyer's and seller's optimal holdings under the principal trade, $a^{P,b}$ and $a^{P,s}$, respectively. Conditional on trading on a principal basis, these two lines define three regions: a customer with assets $a < a^{P,b}$ would be a buyer, with holdings $a > a^{P,s}$ would be a seller, and with current assets $a \in [a^{P,b}, a^{P,s}]$ would not trade on principal. These three regions are a direct consequence of the inclusion of inventory costs. On the one hand, in the principal mechanism, buyers trade at an effective price higher than the one received by sellers. Hence, conditional on preference type, buyers' optimal quantity is smaller than that of sellers. On the other hand, the principal trade surplus of those customers with current holdings between the buyer's and seller's optimal holdings is

Figure 1.3: Optimal asset holdings and trading mechanism choice.



Note: This figure depicts the estimated model policy functions of each customer, conditional on her preference type and current holdings. The lower and upper solid lines represent the buyer's and seller's optimal asset holdings under the principal trade, $a^{P,b}$ and $a^{P,s}$, respectively. The dashed line represents the optimal asset holdings under the agency trade, a^A. Regarding the mechanism choice, the principal and agency regions are shaded in orange and blue, respectively.

smaller than the principal costs faced by the dealers. Hence, there are no gains from trade and those customers decide not to trade on a principal basis. The agency optimal holdings, in turn, are represented by the dashed black line a^A. These positions are between those of the principal buyers and the principal sellers. Recall that agency trading does not imply any cost for dealers. Since dealers face no costs, the transaction cost charged to customers, conditional on trading volume, is smaller. The direct consequence is that the effective agency price is between the effective principal buy and sell prices, and thus agency optimal holdings are between those of the principal traders.

Figure 1.3 also presents the trading mechanism each customer chooses. The blue shaded area represents the agency region: customers who decided to wait for execution instead of paying the cost for immediacy or waiting to contact another dealer. As can be seen, in the estimated model (see section 1.5), every potential principal non-trader, i.e., customers with holdings $a \in [a^{P,b}, a^{P,s}]$, finds that engaging in an agency trade is better than not trading at all and waiting for a new contact with a dealer. Finally, the orange shaded area stands for customers that trade on principal.

To better understand these policy functions, consider for example customers with preferences $\epsilon_i = 0.4$. When contacted by a dealer, these customers compute their optimal asset position as principal traders, $a_i^{P,b}$ or $a_i^{P,s}$, and their expected optimal position after the waiting period of the agency trade, $\bar{a}_i^A$. Given these optimal asset positions, they evaluate, using Equation (1.8), which trade to perform. As Figure 1.3 shows, customers owning roughly less than 0.94 units of the asset perform a principal buy. Customers holding between 0.94 and 1.02 units perform an agency trade. Finally, customers holding assets above 1.02 choose to sell on a principal basis.

Figure 1.3 confirms an earlier observation: principal traders are concentrated in the extremes of the preference-assets state space. Firstly, conditional on preference types, principal trading is mostly performed by customers with current asset holdings far away from their optimal ones. As it was discussed in subsection 1.2.2, this is because the utility loss of each extra unit away from the optimal position is marginally increasing, whereas the principal premium that needs to be paid to avoid such costs is constant. Secondly, conditional on current asset holdings, agency trading is mostly performed by customers with preferences close to the mean. This is because optimal asset positions are increasing in preference types: customers with extreme preferences will find themselves more often far away from their optimal position than customers with moderate preferences. Given the relation between trading mechanism choice and the distance between the current and optimal position, the model tells us that customers with moderate preferences are more likely to perform agency trades, while customers with extreme preferences are more likely to trade on principal.

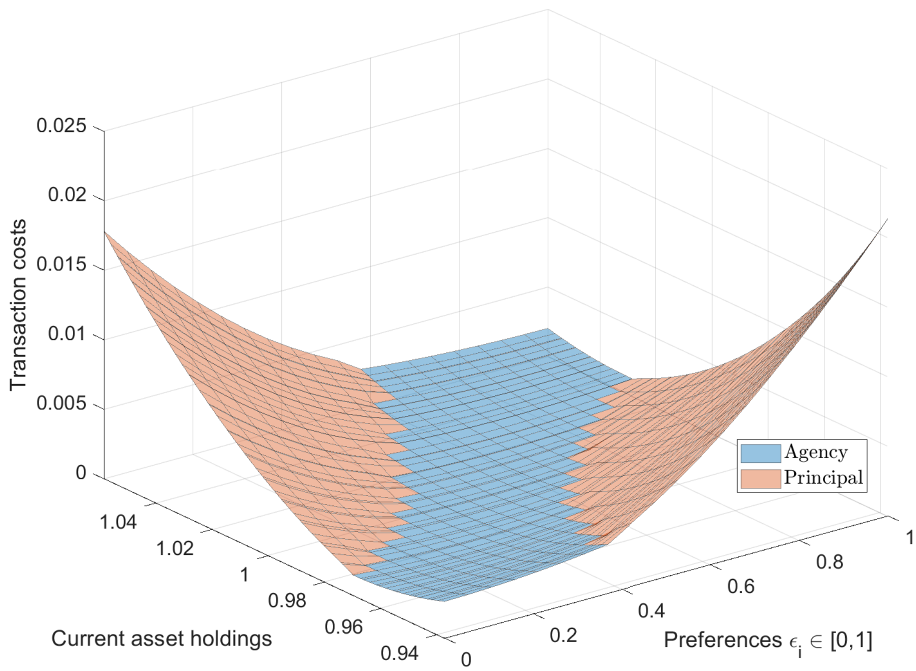
1.3.2 Equilibrium Transaction Costs

I next present the distribution of transaction costs paid by customers. As Equations (1.3) and (1.5) show, these costs are solved through Nash bargaining; therefore, they incorporate the specific characteristics of the trade. Particularly, transaction costs are convex combinations between customers' expected trading surplus and dealers' inventory cost. In turn, these objects are functions of the asset holdings and preference held by the customer when she contacts the dealer, and of the resulting trading mechanism chosen. Figure 1.4, which maps transaction costs with the asset-preference state-space, depicts such heterogeneity.

Overall, the broad features of transaction costs in LR09 still hold. For example, marginal transaction costs are increasing in the traded volume. A marginally decreasing utility implies that, given a certain optimal position the marginal trading surplus is increasing in the volume traded. The bargaining protocol used implies that transaction costs are linear functions of such surpluses; thus, they inherit the property [16]. On top of this, two interesting properties regarding the trading mechanism distinction are observed. Firstly, principal transaction costs are on average larger than those of agency trades. On the one hand, principal traders exchange larger quantities and thus obtain larger trade surpluses. On the other hand, even conditioning on the customer's trading surplus, principal transaction costs are still larger than agency, given the inclusion of the translated inventory costs. This

[16] Pinter et al. (2024) study the relation between trading costs and trading size in the UK government and corporate bond markets. In contrast with other empirical papers on the topic, their database has both customers' and dealers' identities. This feature allows them to control for customer cross-section variation when computing the trade size effect. In line with the model here developed, they show that, conditional on the customer's identity, trading costs are increasing in trade size.

Figure 1.4: Transaction costs under each trading mechanism.



Note: This figure depicts the estimated model transaction cost paid by each customer, conditional on her preference type and current asset holdings. The orange-shaded area refers to principal costs. The blue shaded area refers to (present valued) agency costs.

later feature is evident from the presence of jumps at the thresholds [17]. Secondly, principal transaction costs increase at a higher rate when moving both towards extreme preferences and towards larger trading quantities. When customers trade on agency, they are subject to preference shocks. This implies that agency customers anticipate that both the utility

[17]If current asset holdings equal asset thresholds, the indifference condition (1.8) indicates that the net trade surplus for any preference type under both mechanisms is the same. At such current asset holdings, from the definition of inventory costs and as long as asset holding thresholds and principal optimal holdings are different, inventory costs will be positive. Given that transaction costs are convex combinations of customers' trade surpluses and dealer costs, at the thresholds principal costs exceed (present valued) agency costs exactly by the inventory costs amount.

$$\phi_i^P(\hat{a}_i) - \theta p|a_i^P - \hat{a}_i| = \hat{\beta}\phi_i^A(\hat{a}_i)$$

This result can be easily obtained combining Equations (1.3), (1.5), and (1.8).

they get from current holdings and the optimal trading volume may change while waiting for execution. Hence, instead of the certain immediate trade surplus given by principal trades, agency customers need to consider an average surplus based on expected preference shocks. Therefore, across the agency region expected trade surpluses, and consequently transaction costs, are relatively flatter [18].

As can be seen, the model yields a rich heterogeneity both across and within trading mechanisms. Customers with large (small) trading needs and holding relatively extreme (moderate) preference types choose principal (agency) trades. Accordingly, those customers trading on principal pay an average higher transaction costs than those trading on agency. Finally, given the possibility of changing preferences while waiting for execution, transaction costs are relatively flatter across the state-space within the agency region. These differences will play a key role when addressing composition effects. If the customers that migrate across trading mechanisms when market conditions change paid different costs than the non-migrating ones, then the samples over which transaction costs pre and post-change are measured will not be comparable. The next section computes average transaction costs as empirical researchers would and develops a strategy to control for such change in samples.

1.4 Liquidity Measures

Recent empirical literature on OTC markets argues that liquidity conditions have changed during the last decade. In particular, researchers document a shift in trading volume, from immediate principal towards delayed agency trades, accompanied by an increase in immediacy costs (Anderson and Stulz, 2017; Schultz, 2017; Bao, O'Hara, and Zhou, 2018; Bessembinder, Jacobsen, Maxwell, and Venkataraman, 2018; Dick-Nielsen and Rossi, 2019; O'Hara

[18] In Appendix 1.A.7 I graph transaction costs per dollar traded. All the features previously mentioned hold if this alternative specification is considered.

and Zhou, 2021; Kargar, Lester, Lindsay, Liu, Weill, and Zúñiga, 2021; Choi, Huh, and Seunghun Shin, 2024). In this section I compute the model's liquidity measures necessary to understand and analyze this phenomenon. Firstly, I compute the turnover rate and average transaction costs that serve as theoretical counterparts of the empirical measures. Secondly, I build counterfactual measures of transaction costs which account for composition effects. By comparing average and counterfactual measures I obtain the size and sign of the bias. These objects are used to revisit how liquidity changes when there are higher regulatory costs or when the speed of execution of agency trades increases.

1.4.1 Turnover and Transaction costs

To compute liquidity measures, it is useful to regroup the optimal trading mechanism sets. Define $P_i \equiv Buy_i^P \cup Sell_i^P$, $A_i \equiv Buy_i^A \cup Sell_i^A \cup NoT_i^A$, and $NT_i \equiv NoT_i^P$, as the sets under which customers of preference i trade on principal, on agency, or do not trade at contact with dealers. The turnover rate is computed as the ratio between the total dealer-customer volume traded per unit of time and the aggregate asset supply. The supply of assets is fixed at A, so I only need to compute the volume. Principal trades are performed by customers who are waiting to contact a dealer and prefer immediate trades, i.e., customers in state $n_{[a,i,\omega_1]}$, where $a \in P_i$. These contacts happen at rate α, and the volume traded in each transaction is $|a_i^P(a) - a|$. In turn, agency trades are performed by customers who had already agreed to conduct such contract and therefore are waiting for its execution. These customers are found in states $n_{[a,i,\omega_2]}$, where $a \in \mathcal{A}^*$. They execute their contracts at rate β, and exchange volume according to $|a_i^A - a|$. The turnover in each mechanism, expressed in percentage points, is:

$$\mathcal{T}^P = 100 \times \frac{1}{A} \alpha \sum_{i \in \mathcal{I}} \sum_{a \in P_i} n_{[a,i,\omega_1]} |a_i^P - a|, \tag{1.15}$$

$$\mathcal{T}^A = 100 \times \frac{1}{A}\beta \sum_{i\in\mathcal{I}} \sum_{a\in\mathcal{A}^*} n_{[a,i,\omega_2]}|a_i^A - a|. \tag{1.16}$$

The aggregated turnover is just the sum of the turnovers in both mechanisms, $\mathcal{T} = \mathcal{T}^P + \mathcal{T}^A$. In a similar fashion, the volume-weighted average transaction costs for each trading mechanism can be computed. To do this, I first compute the transaction cost per (*numeraire*) dollar traded. Then these figures are averaged using the total volume share of each contract as weights. A consideration must be made regarding the computation of per-dollar costs for agency trades. In such contracts, transaction costs are arranged at contact with dealers and the optimal asset positions are chosen at execution. While waiting for execution, customers can suffer preference shocks. Hence, two customers with the same agency contract might end up trading different volumes. Hence, I compute the aggregated volume for each contract. To do so, I rely on the Law of Large Numbers and track customers across the state-space while they are waiting for execution. The weighted average transaction cost in each mechanism, expressed in basis points (bps), is:

$$\mathcal{S}^P = 10000 \times \sum_{i\in\mathcal{I}} \sum_{a\in P_i} \frac{n_{[a,i,\omega_1]}|a_i^P - a|}{\sum_{i\in\mathcal{I}} \sum_{a\in P_i} n_{[a,i,\omega_1]}|a_i^P - a|} \frac{\phi_{a,i}^P}{|a_i^P - a|p}, \tag{1.17}$$

$$\mathcal{S}^A = 10000 \times \sum_{i\in\mathcal{I}} \sum_{a\in A_i} \frac{n_{[a,i,\omega_1]} rav_{a,i}}{\sum_{i\in\mathcal{I}} \sum_{a\in A_i} n_{[a,i,\omega_1]} rav_{a,i}} \frac{\phi_{a,i}^A}{rav_{a,i}p}. \tag{1.18}$$

where $rav_{a,i}$ stands for the realized agency volume for contracts signed by customers holding i preference and a assets at the moment of contact with dealers[19]:

[19]Note that $rav_{a,i}$ takes into account the possibility of contracting an agency trade but ending up not trading. This happens whenever the current and optimal asset holdings are equal at execution. An alternative computation tracking agency customer until execution yields the same result.

$$rav_{a,i} =(1-\hat{\delta})|a_i^A - a| + \hat{\delta}\sum_{j\in\mathcal{I}} \pi_j |a_j^A - a|.$$

The average transaction cost unconditional on trading mechanism is just the weighted average of the previous figures: $\mathcal{S} = [\mathcal{T}^P\mathcal{S}^\mathcal{P} + \mathcal{T}^A\mathcal{S}^\mathcal{A}]/\mathcal{T}$. As can be seen, average transaction costs are functions of both the costs associated with each transaction and the steady-state mass of customers who endogenously trade in each mechanism. When the economy changes, these two vectors are affected. Thus, the model is able to capture not only the change in transaction cost per trade, but also the sample composition effects.

1.4.2 Transaction Costs Decomposition

To account for composition effects, I build counterfactual measures of average transaction costs fixing the samples over which they are measured. In order to do so, I decompose the steady-state distribution into those customers that, under alternative parametrizations, would migrate across mechanisms and those that would not. Counterfactual transaction cost measures are computed using only the subsamples of non-migrating customers.

Recall that, when customers contact dealers, they choose their optimal trading mechanism according to thresholds that satisfy the indifference condition (1.8). These thresholds define trading mechanism sets, i.e., preference-specific asset holding sets under which customers choose to trade on principal, on agency, or not to trade at all, P_i, A_i and NT_i, respectively. Consider firstly alternative parametrizations, denoted by q, and compute their steady-state trading mechanism sets. Secondly, for each preference type, compute the intersections across parametrizations between these trading mechanism sets. To ease the exposition, I only consider two parametrizations, $q \in \{0, 1\}$, but the method can be easily extended to account for any number of parametrizations. Table 1.1 presents the resulting subsets. Diagonal cells include customers that choose the same trading mechanism under the two scenarios. I call these customers non-migrants. Conversely, non-diagonal cells include customers who change their optimal mechanism when facing different scenarios. I call these customers migrants. For example, the population of customers with preference i holding assets $a \in P_i^0 \cap A_i^1$ would trade on principal under $q = 0$ and would migrate towards agency under $q = 1$ [20]. These subsets allow defining subsamples over which to compute transaction

Table 1.1: Sample decomposition

	P_i^1	A_i^1	NT_i^1
P_i^0	$P_i^0 \cap P_i^1$	$P_i^0 \cap A_i^1$	$P_i^0 \cap NT_i^1$
A_i^0	$A_i^0 \cap P_i^1$	$A_i^0 \cap A_i^1$	$A_i^0 \cap NT_i^1$
NT_i^0	$NT_i^0 \cap P_i^1$	$NT_i^0 \cap A_i^1$	$NT_i^0 \cap NT_i^1$

costs. To this end, I add new notation. Superscripts attached to cost measures indicate both the trading mechanism and the parameters used. In turn, subscripts, whenever present, denote which trading subsets were used to define the subsample. For example, $\mathcal{S}_{P^0,P^1}^{P,0}$ refers to principal transaction costs paid under scenario $q = 0$ by customers who trade on principal both under $q = 0$ and $q = 1$. In turn, $w_{P^0,P^1}^{P,0}$ refers to the volume share accounted for such transactions under scenario $q = 0$. Finally, I can decompose the change in transaction costs for each mechanism due to a parametric change. Consider $q = 0$ as the initial scenario, and $q = 1$ as the new one.[21]

[20]If $Q > 2$ number of parametrizations are considered, 3^Q number of subsets within a Q-dimension matrix are obtained. The diagonal of such higher-order matrix defines customers that choose the same trading mechanism under all the alternative parametrizations. For example, customers with preference i that remain trading on principal regardless of the parametrization used are those with assets $a \in \cap_{q=1}^{Q} P_i^q$.

[21]See Appendix 1.A.8 for details.

$$
\begin{aligned}
\Delta\mathcal{S}^{P} = \mathcal{S}^{P,1} - \mathcal{S}^{P,0} = & \underbrace{\mathcal{S}^{P,1}_{P^0,P^1} \times w^{P,1}_{P^0,P^1} - \mathcal{S}^{P,0}_{P^0,P^1} \times w^{P,0}_{P^0,P^1}}_{\text{Principal non-migrants}} \\
& + \underbrace{\mathcal{S}^{P,1}_{A^0,P^1} \times w^{P,1}_{A^0,P^1} + \mathcal{S}^{P,1}_{NT^0,P^1} \times w^{P,1}_{NT^0,P^1}}_{\text{Inflow migration}} \\
& - \underbrace{\mathcal{S}^{P,0}_{P^0,A^1} \times w^{P,0}_{P^0,A^1} - \mathcal{S}^{P,0}_{P^0,NT^1} \times w^{P,0}_{P^0,NT^1}}_{\text{Outflow migration}},
\end{aligned} \tag{1.19}
$$

$$
\begin{aligned}
\Delta\mathcal{S}^{A} = \mathcal{S}^{A,1} - \mathcal{S}^{A,0} = & \underbrace{\mathcal{S}^{A,1}_{A^0,A^1} \times w^{A,1}_{A^0,A^1} - \mathcal{S}^{A,0}_{A^0,A^1} \times w^{A,0}_{A^0,A^1}}_{\text{Agency non-migrants}} \\
& + \underbrace{\mathcal{S}^{A,1}_{P^0,A^1} \times w^{A,1}_{P^0,A^1} + \mathcal{S}^{A,1}_{NT^0,A^1} \times w^{A,1}_{NT^0,A^1}}_{\text{Inflow migration}} \\
& - \underbrace{\mathcal{S}^{A,0}_{A^0,P^1} \times w^{A,0}_{A^0,P^1} - \mathcal{S}^{A,0}_{A^0,NT^1} \times w^{A,0}_{A^0,NT^1}}_{\text{Outflow migration}}.
\end{aligned} \tag{1.20}
$$

The introduced decomposition highlights the interaction between the changing average costs in each subsample and the changing subsample weights. It has three components. The first term accounts for the non-migrants' effect. On the one hand, customers who keep on trading under the same mechanism may pay different costs. On the other hand, the volume share of those customers may also change. The second and third terms are related to the migrants' effect. Under a new scenario, some customers may decide to change their optimal trading strategy. Customers that represent an inflow into a given mechanism add up their costs to the overall average. Conversely, customers that imply an outflow subtract their previously paid costs from that average.

Equations (1.19) and (1.20) provide a natural way of defining counterfactual measures of transaction costs free of composition effects. If the samples within the trading mechanism were held constant, non-migrant customers would have full weight in all scenarios. Therefore, I define the composition-free measures of transaction cost under parametrization q, $\tilde{\mathcal{S}}^P(q)$ and $\tilde{\mathcal{S}}^A(q)$, as the costs measured within the non-migrant samples. In turn, the composition-free measures of transaction cost change, $\Delta\tilde{\mathcal{S}}^P$ and $\Delta\tilde{\mathcal{S}}^A$, are set to account only for such non-migrant figures. Finally, the composition effect bias measures, CE^P and CE^A, are defined as the fraction of the change in transaction costs due to migration.

$$\tilde{\mathcal{S}}^P(q) \equiv \mathcal{S}^{P,q}_{P^0,P^1}, \tag{1.21}$$

$$\tilde{\mathcal{S}}^A(q) \equiv \mathcal{S}^{A,q}_{A^0,A^1}, \tag{1.22}$$

$$\Delta\tilde{\mathcal{S}}^P \equiv \mathcal{S}^{P,1}_{P^0,P^1} - \mathcal{S}^{P,0}_{P^0,P^1}, \tag{1.23}$$

$$\Delta\tilde{\mathcal{S}}^A \equiv \mathcal{S}^{A,1}_{A^0,A^1} - \mathcal{S}^{A,0}_{A^0,A^1}, \tag{1.24}$$

$$CE^P \equiv 1 - \Delta\tilde{\mathcal{S}}^P/\Delta\mathcal{S}^P, \tag{1.25}$$

$$CE^A \equiv 1 - \Delta\tilde{\mathcal{S}}^A/\Delta\mathcal{S}^A. \tag{1.26}$$

The introduction of composition-free measures of transaction cost changes sheds light on the necessary conditions for the existence of composition effects mentioned in the introduction of this paper. In the first place, migrating customers are needed. Their absence would imply that the samples under the two scenarios are equal. Secondly, the costs paid by migrating and non-migrating customers should be different. Otherwise, the in-flowing and out-flowing migrants would not alter the average costs of each mechanism. Finally, as long as the difference between costs paid by migrants and non-migrants is driven by unobservable characteristics, empirical estimates would include a composition effect bias. Our model suggests that such an unobservable characteristic is the idiosyncratic trading surplus of each customer, which in turn is a function of both the distance between current and optimal positions and the idiosyncratic utility each customer derives from holding the assets.

1.5 Estimation

In this section I bring the model to the data. Particularly, I target key moments of the US corporate bond secondary market. I initially outline the estimation method. Later I describe how to compute the moments used in such a procedure, both theoretically and empirically. Finally, I present the estimation results and the moments' variation that allows for the identification of the parameters.

1.5.1 Estimation Procedure

The baseline parametrization of the model will consist of a combination of externally calibrated parameters and estimated parameters. I set the unit of time to be a month. In line with recent research on structural estimation of related search models (Coen and Coen, 2022; Pinter and Uslu, 2022), I consider a monthly discount rate of 0.5%. The support of the preferences shifters ϵ_i is normalized to $\left\{\frac{i-1}{I-1}\right\}_{i=1}^{I}$, with $I = 20$. In the model, expanding or contracting the support of ϵ_i only scales up or down the nominal variables, i.e., the inter-dealer price and the transaction costs. Given that I will focus on transaction costs per (*numeraire*) dollar traded, normalizing such support does not affect the results. Similarly, the supply of assets A only scales up and down both nominal and real variables. Since all real variables will be expressed in terms of the total asset supply, I normalize $A = 1$. As was shown in subsection 1.2.2, the bargaining power of the dealers, η, is closely related to the arrival rate of opportunities to trade, α. In a nutshell, customers are indifferent between contacting high bargaining power dealers often and low bargaining power dealers scarcely. This precludes me from disentangling these two parameters, and therefore I opt to externally calibrate the bargaining power and to estimate the contact rate with dealers. I follow Hugonnier, Lester, and Weill (2020) and set $\eta = 0.95$. Finally, the last object externally calibrated is the probability distribution assigned to each preference type. I follow Coen and Coen (2022) and assume that such preferences are uniformly distributed, $\pi_i = 1/I \ \forall i$. In the appendix 1.C.2 I show that the main results qualitatively hold when considering lower or higher bargaining powers or alternative preference distributions.

The remaining parameters of the model are the rates at which customers contact dealers, suffer preference shocks and execute their agency trades, α, δ and β respectively, the dealer's marginal inventory costs, θ, and the utility curvature parameter, σ. I jointly estimate these parameters using the generalized method of moments (GMM). Particularly, I define the vector $\upsilon = [\alpha, \delta, \beta, \theta, \sigma]$ and estimate $\hat{\upsilon}$ as the argument that minimizes the percentage difference between the implied theoretical moments, $m(\upsilon)$, and the computed empirical moments, m_s:

$$\hat{\upsilon} = \underset{\upsilon \in \Upsilon}{\arg\min}[(m(\upsilon) - m_s) \oslash m_s]' W [(m(\upsilon) - m_s) \oslash m_s],$$

where $\oslash$ is element-wise division. Note that by using percentage deviation I ensure that the scales of the different moments do not play any role in the procedure. In line with the literature, W is set as the identity matrix, thus assigning equal weights to the different moments (Coen and Coen, 2022; Pinter and Uslu, 2022).

1.5.2 Moments

I choose a set of moments that covers both quantities and prices, as well as the interaction among them. I target the overall monthly turnover, $\mathcal{T}$, the volume weighted average transaction costs in each mechanism, S^P and S^A, and the slopes of the transaction costs over the trade size, for each mechanism, γ^P and γ^A. In particular, to gauge the size of composition effects, it is fundamental to target the differential transaction costs paid by migrants and non-migrants. Section 1.6 will show that migrants are located in the extremes of the trading size distribution, conditional on preference type. Thus matching the slope of transaction costs on trading size, γ^P and γ^A, informs about the differential transaction costs paid by migrants and non-migrants. In subsection 1.5.3 I discuss how the variation of these moments can identify the vector of parameters υ.

Theoretical Moments

For any given vector υ, I compute the theoretical moments using the steady-state equilibrium of the model. These are:

- Monthly turnover:

$$\mathcal{T} = 100 \times \frac{\alpha \sum_{i\in\mathcal{I}} \sum_{a\in P_i} n_{[a,i,\omega_1]} |a_i^P - a| + \beta \sum_{i\in\mathcal{I}} \sum_{a\in\mathcal{A}^*} n_{[a,i,\omega_2]} |a_i^A - a|}{A}, \tag{M.1}$$

- Volume weighted average transaction cost in each mechanism:

$$\mathcal{S}^P = 10000 \times \sum_{i\in\mathcal{I}} \sum_{a\in P_i} \frac{n_{[a,i,\omega_1]} |a_i^P - a|}{\sum_{i\in\mathcal{I}} \sum_{a\in P_i} n_{[a,i,\omega_1]} |a_i^P - a|} \frac{\phi_{a,i}^P}{|a_i^P - a| p}, \tag{M.2}$$

$$\mathcal{S}^A = 10000 \times \sum_{i\in\mathcal{I}} \sum_{a\in A_i} \frac{n_{[a,i,\omega_1]} rav_{a,i}}{\sum_{i\in\mathcal{I}} \sum_{a\in A_i} n_{[a,i,\omega_1]} rav_{a,i}} \frac{\phi_{a,i}^A}{rav_{a,i} p}. \tag{M.3}$$

- Transaction cost - trade size slope in each mechanism:

$$\gamma^P = 100 \times \frac{cov(\phi^P / (|a^P - a| p), |a^P - a|)}{var(|a^P - a|)}, \tag{M.4}$$

$$\gamma^A = 100 \times \frac{cov(\phi^A / (rav \times p), rav)}{var(rav)} \tag{M.5}$$

where the variance and covariance equations are described in the Appendix 1.B.1.

Empirical Moments

To compute the empirical moments, I rely on transaction data of the US corporate bond secondary market, from January 2016 to December 2019. Specifically, I use the academic Trade Reporting and Compliance Engine (TRACE) database, produced by the Financial Industry Regulatory Authority (FINRA).

Given the well-known presence of reporting errors, the data is filtered following the procedure outlined in Dick-Nielsen and Poulsen (2019) [22]. I also remove the duplicated inter-dealer trades and those trades in which dealers transfer bonds to their non-FINRA affiliates for book-keeping purposes (Adrian, Boyarchenko, and Shachar, 2017) [23]. I further merge this transaction-level data with bond-level variables from the Mergent Fixed Income Securities Database (FISD). Following the empirical literature, several filters are applied (e.g., Bessembinder, Jacobsen, Maxwell, and Venkataraman, 2018; Friewald and Nagler, 2019; Kargar, Lester, Lindsay, Liu, Weill, and Zúñiga, 2021). Among them, the most significant are dropping bonds that are preferred, convertible or exchangeable, yankee bonds, bonds with a sinking fund provision, variable coupon, with time to maturity of less than a year, or issued less than two months before the transaction date[24].

Needless to say, the empirical transaction costs are partially driven by features not present in my model, e.g., default risk and asymmetric information. In that regard, to improve the likelihood of my model capturing the targeted moments I exclude from the sample those bonds that had been labeled as high yield at any point during my sample period [25].

[22]Both the algorithm and the filter results can be downloaded from my personal website.

[23]Starting on November 2, 2015, FINRA provides explicit labels for the so-called book-keeping trades.

[24]I also remove bonds that are security backed, equity-linked, putable, foreign-currency denominated, privately placed, perpetual, sold as part of a unit deal, or secured lease obligations bonds.

[25]Using standard letter-number equivalences (e.g., AAA=1, D=25), I average the letter ratings of the three agencies present in FISD: S&P, Moodie's and Fitch. I then go back to letter ratings using the same equivalence and classify as high yield a bond with a rating equal to or lower than BB+.

One important feature of the academic version of TRACE is that it contains anonymous identities for each dealer. I exploit that feature to identify principal and agency trades. The idea underlying the identification is that the shorter the time it takes for a dealer to offload a position, the bigger it is the probability that those trades had been previously arranged and thus intermediated on an agency basis (Bessembinder, Jacobsen, Maxwell, and Venkataraman, 2018; Kargar, Lester, Lindsay, Liu, Weill, and Zúñiga, 2021; Choi, Huh, and Seunghun Shin, 2024). I classify customer-dealer trades into three categories: those that are quickly offset with other customers, those that are quickly offset with other dealers, and those that are not offset. The first and third categories are agency and principal trades, respectively. Specifically, for each customer-dealer trade, I look for all the offsetting trades of the same dealer in the same bond, within a 15-minute window. If at least 50% of its volume was offset, and the majority of such volume was offset with customers, I label it as an agency trade. If less than 50% of its volume was offset, I label it as a principal trade. The remaining transactions are disregarded.

Two subtleties about the principal/agency distinction are worth noting. First, this procedure allows for multiple matching, in the sense that a single trade can be offset by several trades of the opposite direction. Second, the algorithm may encounter competing trades. In such case I form pairs with the trades that are closer in time firstly, and closer in volume secondly [26].

Once the data has been filtered and only principal or agency customer-dealer trades are kept, I proceed to compute the empirical moments. Turnover and average transaction

[26] Consider for example a dealer that performed four trades in a day, all of them with customers. In trade A the dealer sells 7K at 10:03 am, in trade B she buys 10K at 10:05 am, in trade C she sells 6K at 10:10 am, and in trade D she sells 3K at 10:10 am. In this case, the trades A, C, and D are competing to match with trade B. First I match by time distance, thus trades A and B form a pair. Trade A offsets all of its volume, so it is considered an agency trade. Trade B offsets 70% of its volume. The remaining 30% of the 10k are left to be matched with trades C and D. Given that these last trades happened at the same time, I match according to volume difference. Hence I form a pair with the remaining 3K of trade B and trade D. Again, trade D offsets all of its volume, so it is labeled as agency. Trade B offsets all of it volume as well, against A and D, so it is labeled as agency as well. In turn, trade C is labeled as principal.

costs are first calculated at the bond level and later summarized using medians. In turn, the slopes of transaction costs over trade size are computed using a unique regression for each mechanism subsample. To remove outliers, the sample of bonds is restricted to those that have at least ten observations for each moment computed. The final sample consists of 2829 securities, which add up 1,602,438 observations. Subscripts t, b, d account for customer-dealer trades of a particular bond and during a specific day, respectively.

- Bond b monthly turnover:

$$\mathcal{T}_b = 100 \times \frac{\sum_t vol_{t,b}/iao_b}{k_b/30.5}, \tag{M.6}$$

where $vol_{t,b}$ is the notional volume in trade t, k_b is the day count after offering and before maturity within the period sample, and iao_b is the average amount outstanding during those k_b days. Note that this specification accounts for months in which the bond has no trades at all.

- Bond b volume-weighted average transaction cost in each mechanism:

$$\mathcal{S}_b^P = \sum_{t,d}(s_{t,b,d} \times vol^P_{t,b,d}) / \sum_{t,d} vol^P_{t,b,d}, \tag{M.7}$$

$$\mathcal{S}_b^A = \sum_{t,d}(s_{t,b,d} \times vol^A_{t,b,d}) / \sum_{t,d} vol^A_{t,b,d}, \tag{M.8}$$

where $s_{t,b,d}$ is Choi, Huh, and Seunghun Shin (2024)'s Spread1:

$$s_{t,b,d} = Q \times 10000 \times (\frac{p_{t,b,d} - p^{DD}_{b,d}}{p^{DD}_{b,d}}) \quad , \quad p^{DD}_{b,d} = \frac{\sum_{t \in DD_{b,d}} vol^{DD}_{b,d,t} p^{DD}_{b,d,t}}{\sum_{t \in DD_{b,d}} vol^{DD}_{b,d,t}}$$

with $Q = 1$ (-1) if a customer buys (sells). To reduce the noise coming from micro trades, I only consider trades in which the volume >\$100K (Pinter, Wang, and Zou, 2024). Since prices are expressed per fixed amount of bond units, the percentage difference between the customer-dealer price $p_{t,b,d}$ and the inter-dealer price $p_{b,d}^{DD}$ equals the transaction costs per dollar computed in the model.

- Transaction cost - trade size slope in each mechanism. I estimate the following model for each mechanism subsample:

$$s_{t,d,b} = \alpha + \beta FE + \gamma 100(vol_{t,b,d}/iao_b) + \epsilon_{t,b,d}, \tag{M.9}$$

  where $FE = [dealer, bond, day]$. Given that in the model the asset supply is normalized, to match the theoretical counterpart I consider the ratio between the volume traded and the amount outstanding. In that regard, the OLS estimates $\hat{\gamma}^P$ and $\hat{\gamma}^A$ are interpreted as how many bps transaction costs increase with a one percentage point increase in the traded amount outstanding of the bond. Appendix 1.B.2 presents the regression results.

1.5.3 Estimation results

Table 1.2 presents the estimation results. To the best of my knowledge, this is the first paper to structurally estimate a search model using the US corporate bond secondary market data. Given this lack of reference, I limit the exposition to explain what the parameter values mean for our model and, when possible, trace comparisons with empirical observations.

The estimation results tell us that search frictions matter. This is not only because customers need to wait a significant amount of time to contact dealers, but also because when they do so, they only partially realize their gains from trade. Customers contact dealers around 9 times per month, which means that they have to wait around 2 business days for an opportunity to update their holdings. In the model, the rate at which customers

Table 1.2: Baseline Calibration.

Parameter	Description	Value
- Normalized -		
	Unit of time	1 month
A	Asset supply	1
ϵ_i	Preference shifter	$\left\{\frac{i-1}{I-1}\right\}_{i=1}^{20}$
- Externally calibrated -		
r	Discount rate	0.5%
π_i	Preference shifter distribution	$1/I$
η	Dealer's bargaining power	0.95
- GMM estimated -		
α	Contact with dealer rate	9.15
δ	Preference shock rate	2.59
β	Agency execution rate	1.00
θ	Inventory cost (bp)	0.89
σ	Utility curvature	2.73

contact dealers is as important as the trade surplus they can preserve after paying transaction costs. In other words, what matters is the bargaining-adjusted rate, $\alpha(1-\eta)$. This later rate tells us that customers need to wait around 2 months to fully extract all the trading surplus from rebalancing positions.

Preference shocks happen with less intensity than trade opportunities. On expectation, a customer changes preferences around 2.5 times per month[27]. On the one hand, the fact that customers change preferences less often than the rate at which they contact dealers means that not all trading opportunities are realized. On the other hand, whenever trading does happen, the amount exchanged is larger than what it would be with a higher preference shock rate: customers can take more extreme positions knowing that those positions will stay optimal longer. In turn, these larger amounts exchanged translate into higher transaction costs. Both infrequent trading and high transaction costs are salient features of the secondary

[27] While preference shocks arrive at a Poisson rate of 2.59, the probability of receiving a preference different from the current one is 95%, given the uniform distribution and a support of 20 types.

corporate bond market.

In comparison with LR09, two parameters are added. The first of them is β, which accounts for the expected execution delay of an agency trade. The available data inform us only of when trades are executed, but not on the initial customer-dealer contact that started the transaction. The estimation results shed light on this scarcely explored parameter and tell us that the execution waiting times are considerable. Dealers take on expectation one month to execute trades for those customers not willing to pay the principal premium.

Regarding the second novel parameter, the marginal inventory costs θ, the results suggest that these are considerable: 0.89 bps for a one-way trade. To interpret this number, let me focus on the regulation-induced costs dealers face when including assets in their inventories. The empirical evidence indicates that the leverage ratio requirement (LRR) is the most tightly binding constraint for most U.S. banks after the post-2008 financial crisis regulations were set (Duffie, 2017; Greenwood, Stein, Hanson, and Sunderam, 2017). The LRR requires banks to hold capital for an amount of 5% of the non-risk-weighted value of assets in inventory[28]. Restricting attention to this most binding regulation, the inventory cost faced by a dealer buying $p(a' - a)$ worth of assets, with an average holding period of 10.6 days (Goldstein and Hotchkiss, 2020) and incurring a daily opportunity cost of $r/30\%$, is $5\%[p(a'-a)(e^{(r/30)10.6}-1)]$. The model counterpart of such round-trip principal trade cost would be $2\theta_{LRR}p(a'-a)$, where θ_{LRR} would consider only this specific but important piece of regulation. The following mapping is obtained: $\theta_{LRR} = 5\%[e^{(r/30)10.6} - 1]/2 = 0.44$bps. The comparison between the estimated marginal inventory costs and this back-of-the-envelope LRR cost indicates that the estimation is in the right order of magnitude, arguably capturing other non-regulatory inventory costs.

Finally, the curvature of the utility function is estimated at 2.73. This parameter

[28]The percentage is 3% for non-global systemically important banks with assets over 250 billion dollars, and 5% for global systemically important banks.

Table 1.3: Model Fit

Moment	Empirical			Theoretical
	p50 (m_s)	p25	p75	$m(\hat{v})$
$\mathcal{S}^P$, Principal Vol Weighted Avg Costs	9.12	5.87	14.20	10.29
$\mathcal{S}^A$, Agency Vol Weighted Avg Costs	5.00	2.56	8.73	4.04
$\mathcal{T}$, Monthly Turnover	3.27	2.28	4.61	3.47
	$\hat{\gamma}$ (m_s)	$\hat{\gamma} - s.e.$	$\hat{\gamma} + s.e.$	
γ^P, Principal Cost-Size slope	1.45	1.33	1.58	1.31
γ^A, Agency Cost-Size slope	0.61	0.50	0.73	0.69

Note: Theoretical moments are computed at the steady state, using the calibration presented in Table 1.2. Empirical volume-weighted average cost and monthly turnover are computed at the bond level and summarized by computing the median and interquartile range. Empirical transaction costs - trade size slope is computed estimating Equation (M.9).

is related both to the intensive margin of trading and to the marginal trading surplus. As preferences approach the linear case, the amounts traded increase, with low (high) preference customers selling (buying) as much as possible. On the other hand, as preferences become linear, the marginal surplus from trading an extra unit becomes constant. The estimated value suggests that when customers rebalance positions, they do so in a moderate way, and that the marginal surplus from trading is increasing.

Table 1.3 presents the comparison between the theoretical moments and the empirical ones. Although existing tensions in the model prevent it from perfectly matching the targeted parameters, the results tell us that the model can fairly represent the stylized facts this paper is interested in.

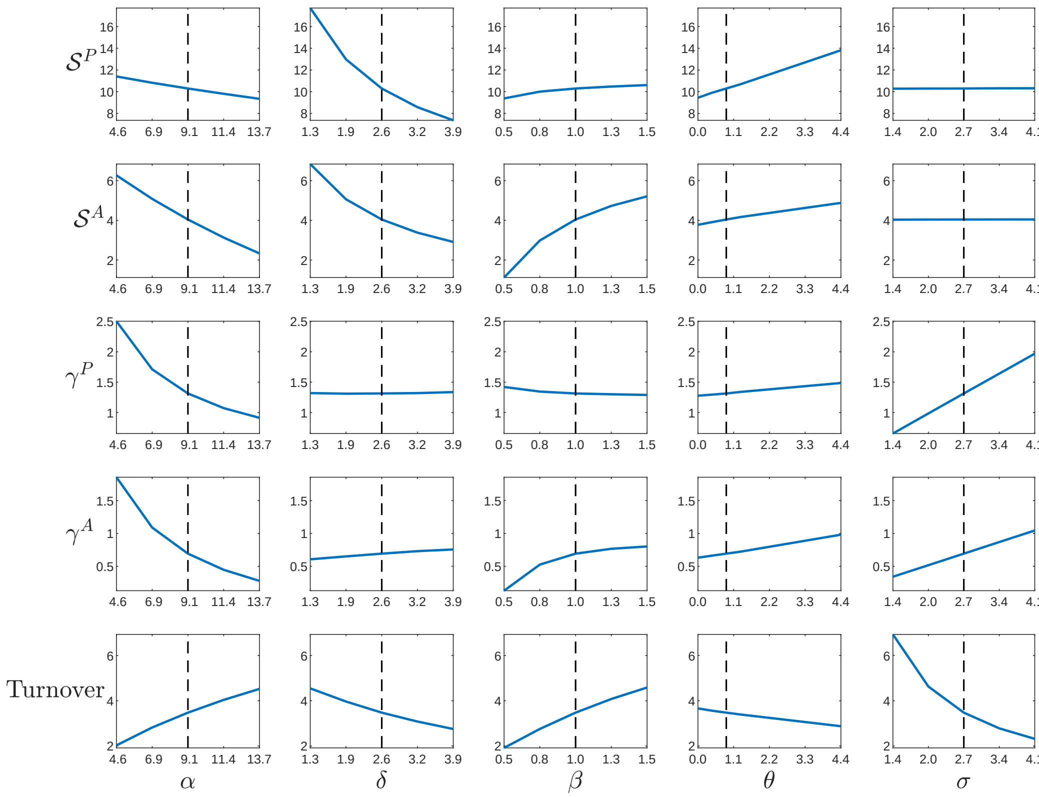
1.5.4 Identification

In this subsection I argue that the moments chosen are informative to jointly pin down the parameter values. In this regard, a common feature in search models of financial markets is the prevalence of general equilibrium effects. Typically, assets are valued according to the utility flow and trading opportunities they generate while customers travel across the state space. Particularly for the model here presented, an asset position would also determine the likelihood of choosing a given trading mechanism. This model structure implies that all parameters affect directly or indirectly the policy functions and correspondingly the observable moments generated. Figure 1.5 shows that, despite these general equilibrium effects, the different directions and intensities in which parameters and moments relate allow me to draw a unique mapping between them.

The first column of Figure 1.5 tells us how the theoretical moments change as we shift the contact rate with dealers. For this, I solve the model for alternative values of α while keeping all other parameters at their estimated values. Not surprisingly, turnover is increasing in the contact rate. The extensive margin increases as more contacts allow customers to trade more often. The intensive margin also increases as optimal asset holdings become more extreme: the expected time of holding unwanted positions is reduced. Perhaps less obvious is the diminishing effect α has on average transaction costs and on cost-size slopes. These figures decrease mainly for the same reason, the surplus from trading is reduced as trading opportunities become more frequent.

The rate at which customers receive preference shocks has the opposite effect on turnover. Although the extensive margin increases – the fraction of customers that contact dealers holding unwanted positions increases – the decrease in the intensive margin dominates. The latter is due to customers opting for less extreme positions, in anticipation of more frequent preference shocks. Regarding transaction costs, as customers expect to change preferences more often, the trading surplus decreases, and so transaction costs decrease as well. The relation between costs and trade size remains mostly unaffected by this parameter. This last (lack of) effect hints at why including both average transaction costs and transaction costs - trade size slopes helps to identify the parameters. For example,

Figure 1.5: Theoretical Moments Variation.



Note: This figure depicts the theoretical moments' variation as parameters change around their estimated values, which are presented with vertical dashed lines. These parameters are the contact with dealers rate, α, the preference shock rate, δ, the agency execution rate, β, the inventory cost expressed in basis points, θ, and the utility curvature, σ. Unchanged parameters are set at their estimated values.

changes in α affect trading size and marginal trading surplus / transaction costs in opposite directions, thus affecting the transaction costs - trade size slopes. Contrastingly, shifts in δ move both in the same direction, without significant effects on the implied slopes.

The third parameter in Figure 1.5 is the execution rate of agency trades, β. Similar to the effect of α on turnover, increasing the execution rate increases the extensive margin of both agency and principal turnover. Although optimal asset positions do not significantly change, migration across mechanisms and other general equilibrium effects (see section 1.6.2) imply that the intensive margin also increases. Consequently, the overall effect on turnover is positive. As expected, a change in the execution rate does not have a major impact on principal transaction costs or on principal cost-size slope. An increase in β makes the agency contract more valuable, given that customers will hold unwanted positions for less time, so both agency average transaction costs and transaction costs derivative on trade size increase. The contrasting effect that the execution rate has on agency and principal related moments is the main source of identification of this parameter.

In turn, an increase in marginal inventory costs θ decreases turnover and increases principal transaction costs, in line with the empirical literature findings (see subsection 1.1.1). Due to general equilibrium effects, agency transaction costs increase as well. Basically, a less dispersed equilibrium asset distribution makes the waiting stage for agency trades less costly. Agency trading surpluses increase and dealers bargain larger transaction costs. This will be explained in detail in section 1.6.1.

Finally, the curvature of the utility function σ, as previously anticipated, plays two main roles. Firstly, as preferences become linear the optimal asset positions become more dispersed and the average trade size becomes larger. This effect increases turnover. Secondly, a lower curvature is translated into lower marginal trade surpluses and hence into lower marginal transaction costs. Therefore costs-size slopes decrease. What distinguishes σ from other estimated parameters, and hence accounts for its main source of identification, is the fact that this parameter does not affect average transaction costs. On the one hand, customers trade larger amounts thus they pay larger transaction costs. On the other hand, conditional on trade size, transaction costs decrease. Overall, these two effects cancel out, resulting in a null effect over average transaction costs.

1.6 Numerical Exercises

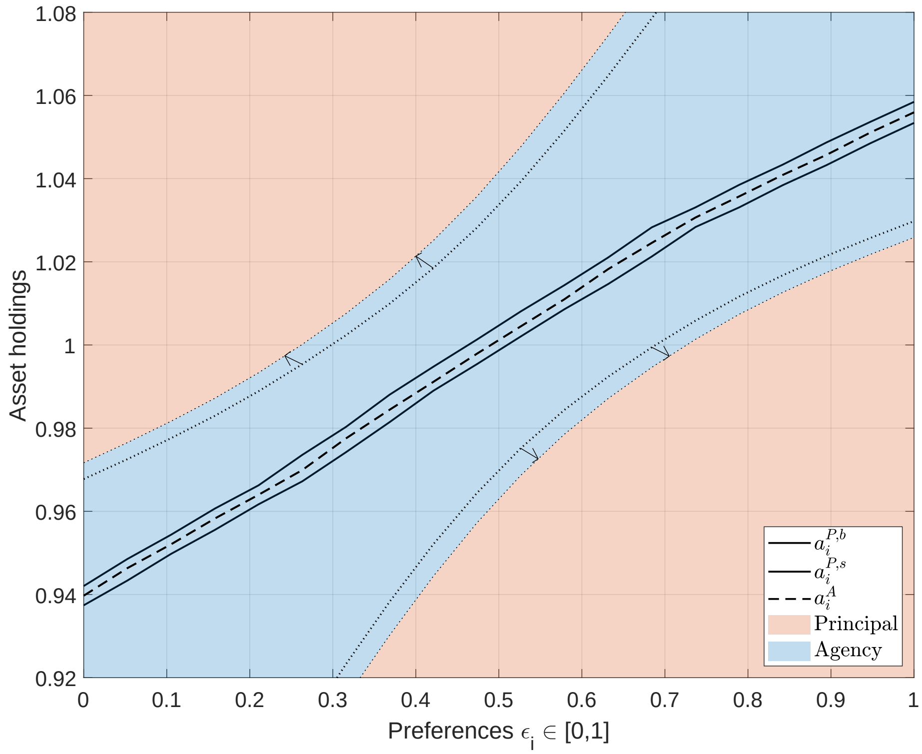
In this section, I use the estimated model to revisit the evidence related to the two major changes observed in the US corporate bond markets in the last decade. First, I address the introduction of post 2008 financial crisis regulations by increasing the models' inventory costs. Second, motivated by the rising popularity of electronic trading venues, I analyze the effects of reducing the execution delay of agency trades. In both cases, when the economy moves through the parametric space, migration across mechanisms appears. Using the proposed decomposition, I show that composition effects account for an economically significant fraction of the changing costs.

1.6.1 Increase in Inventory Costs

An extensive empirical literature has shown how the stricter regulations implemented in the aftermath of the 2008 financial crisis increased dealers' inventory costs, raised the cost of principal trades and shifted volume towards larger agency intermediation (Anderson and Stulz, 2017; Schultz, 2017; Bao, O'Hara, and Zhou, 2018; Bessembinder, Jacobsen, Maxwell, and Venkataraman, 2018; Dick-Nielsen and Rossi, 2019; Choi, Huh, and Seunghun Shin, 2024). Here I revisit such evidence using the tools previously developed. I initially set the inventory costs to a smaller value, $\theta = 0.1$ bps, and then I increase it towards the estimated one. Figure 1.6 shows the policy functions change as we increase inventory costs.

An increase in dealers' inventory costs makes principal trades more expensive. As a consequence, customers migrate towards agency trading. To highlight such migration, Figure 1.6 includes the low inventory cost case thresholds as dotted lines within the baseline calibration agency region. As can be seen, the agency region expands, being the migrating

Figure 1.6: Policy functions as inventory costs increase.



Note: This figure depicts the policy functions of each customer, conditional on her preference type and current holdings, considering $\theta = 0.89$ bps. The lower and upper solid lines represent the buyer's and seller's optimal asset holdings under the principal trade, $a^{P,b}$ and $a^{P,s}$, respectively. The dashed line represents the optimal asset holdings under the agency trade, a^A. Regarding the mechanism choice, the principal and agency regions are shaded in orange and blue, respectively. To ease the comparison across calibrations, the trading mechanism thresholds under $\theta = 0.1$ bps, are depicted as dotted lines within the agency region, and the arrows denote its expansion.

customers those with smaller trading needs [29].

Figure 1.7 presents the liquidity measures computed for $\theta \in [0.1\text{bps}, 0.89\text{bps}]$. Panel A shows that, as inventory costs increase, the overall turnover (black solid line) decreases. This is due to the combination of both extensive and intensive margins going in the same direction. On the one hand, fewer principal trades are being performed, due to the migration towards

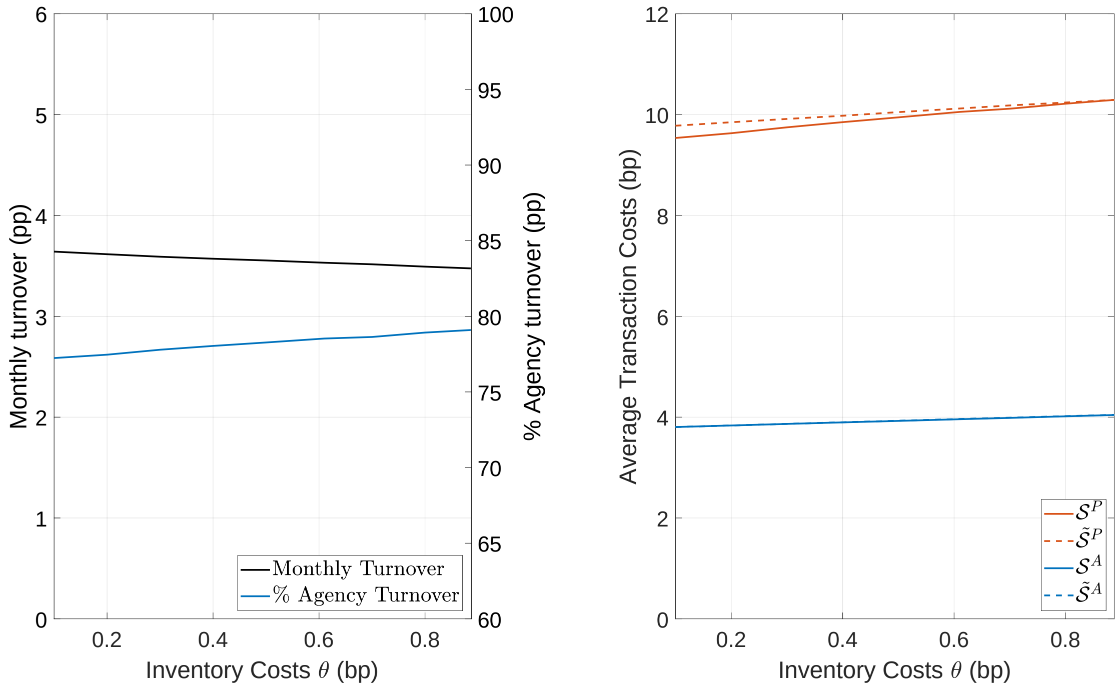
[29]The optimal asset positions are also affected by an inventory costs increase. Such change is depicted in Appendix 1.C.1. Since principal trading becomes more expensive, the trade size decreases: buyers (sellers) have lower (higher) optimal asset positions.

agency. Given the delayed execution of agency trades, overall daily trading decreases. On the other hand, the larger effective prices of principal trading make the average volume per trade decrease in such a mechanism and in the entire distribution. As expected, a positive relation between inventory costs and agency share (blue solid line) is present, which is explained by the aforementioned migration of trades.

Transaction costs are jointly determined with trading volumes. Panel B presents the average costs for each mechanism, $\mathcal{S}^P$ and $\mathcal{S}^A$, in solid lines. As inventory costs rise, dealers translate a fraction of such increase through higher transaction costs, and so principal trading costs mechanically rise. Comparing the two extremes of the inventory costs range considered, the average principal cost increases by $\Delta\mathcal{S}^P = 0.76$ bps. Even though agency trades are not directly related to inventory costs, the transaction cost of these trades increase as well, by $\Delta\mathcal{S}^A = 0.239$ bps. As was previously explained, the effect of inventory costs on agency costs is due to a general equilibrium effect. Given that fast trading becomes more costly, customers expect to hold their positions for longer. Therefore, when choosing these positions, they do so more moderately and the asset dispersion shrinks (see figure 1.C.1). This implies that the burden of holding unwanted positions during the agency trade decreases, increasing both the agency surplus and its transaction costs.

The correlations between inventory costs, migration across mechanisms and average transaction costs have been broadly documented by both the empirical and the theoretical literature. Contrastingly, the self-selection of such migration and the consequent composition effect on cost measures has been largely overlooked. Panel B of Figure 1.7 accounts for such composition effects using the proposed decomposition. I use dashed lines to plot the counterfactual composition-free measures, $\tilde{\mathcal{S}}^P$ and $\tilde{\mathcal{S}}^A$, for each trading mechanism. The comparison of average and counterfactual measures allows us to gauge the sign and size of the bias.

Figure 1.7: Liquidity measures as inventory costs increase.



Note: Panel A (left) presents the steady-state total daily turnover rate, $\mathcal{T}$, and the agency percentage of such figure, $\mathcal{T}^{\mathcal{A}}/\mathcal{T}$, across $\theta \in [0.1\text{bps}, 0.89\text{bps}]$. Panel B (right) presents the steady-state volume-weighted average transaction costs for both mechanisms across $\theta \in [0.1\text{bps}, 0.89\text{bps}]$. Solid lines represent the average measures, $\mathcal{S}^{\mathcal{P}}$ and $\mathcal{S}^{\mathcal{A}}$, whereas dashed lines represent the counterfactual composition-free measures, $\tilde{\mathcal{S}}^{\mathcal{P}}$ and $\tilde{\mathcal{S}}^{\mathcal{A}}$.

Let me start by addressing principal costs. The migration pattern presented in Figure 1.6 tells us that principal customers can be split into non-migrants and outflowing migrants. When marginal inventory costs are set at $\theta = 0.1$ bps, the composition-free measure, i.e., the transaction cost paid by non-migrants, is already 0.24 bps larger than the mechanism's average. Such difference is understood going back to Figure 1.6, where it is observed that non-migrant principals are customers with relatively more extreme preferences and more extreme asset positions, both characteristics associated with higher transaction cost payments. As inventory costs increase, some customers migrate towards agency trading and the proportion of non-migrants increases. This process happens until the entire principal sample is composed by non-migrants. Mechanically, at the highest inventory cost considered, the composition-free and the average measures are equal. Therefore, the change in composition-free transaction costs is smaller than that of the mechanism's average: $\Delta\tilde{\mathcal{S}}^P = 0.51$bp. The difference is explained by a composition effect of $CE^P = 32.2\%$ [30]. In other words, when inventory costs increase, the average willingness to pay of the resulting sample increases, given that those customers who remain trading on principal are the ones who had a higher willingness to pay before costs increased. Therefore, the average transaction cost change captures this increase in the average willingness to pay, and is consequently biased upwards.

Regarding agency trades, the migration pattern associated with increasing inventory costs tells us that customers in this mechanism can be separated into non-migrants and inflowing migrants. At $\theta = 0.1$ bps, the entire agency sample is composed by non-migrants. Therefore, at such parametrizations composition-free and average costs are equal. As inventory costs increase, principal traders migrate towards agency, building up the proportion of inflowing migrants within the agency sample. At the highest inventory costs considered, I find that agency non-migrants pay only 0.07% higher costs than the mechanisms' average. This mild difference contrasts with the principal case, and it is explained by the small transaction costs dispersion found within agency customers, which implies that inflowing migrants pay similar costs to non-migrant customers (see Figure 1.A.1). Given this similarity, composition effects are not expected to play an important role in agency transaction cost measures. As a matter of fact, when comparing the two extremes of the parametric range considered, the composition-free measure equals $\Delta\tilde{\mathcal{S}}^A = 0.242$bp, only 0.003bp above $\Delta\mathcal{S}^A$. Correspondingly, for the agency case, I find a mild composition effect bias of $CE^A = -1.2\%$.

To sum up, the model's predictions are in line with both the empirical and the theoretical literature that studies the effects of raising the intermediaries' inventory costs. In a nutshell, the provision of inventory-related services becomes more expensive, and interme-

[30]Whenever $\tilde{\mathcal{S}}^P$ and $\mathcal{S}^P$ are linear on θ, the composition effect bias is constant. Figure 1.7 indicates that the computed slopes can be well approximated by linear functions.

diation shifts away from principal towards agency trading. Nevertheless, the exercise also shows that transaction cost measures should be revisited, considering the impact that composition effects may have on them. Specifically, I find that these effects account for around a third of the increase in principal costs, and for a negligible figure on agency cost increases.

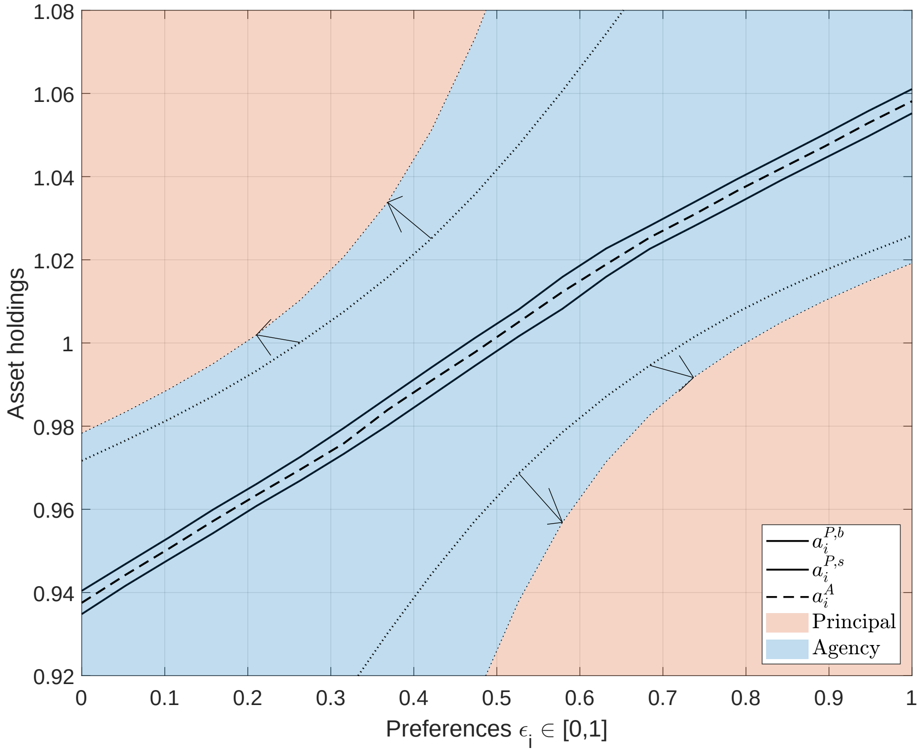
1.6.2 Decrease in the Execution Delay

Corporate bonds have been traditionally traded via voice messages. However, electronic platforms in which customers and dealers can contact counterparties simultaneously have gained popularity in the last decade. Not surprisingly, the empirical research shows that the increase in electronic trading made it easier for dealers to match counterparties in agency trades. Not only the agency share is higher for those bonds that are traded electronically, but also dealers use electronic platforms to find suitable counterparties for customers that contacted them through traditional voice messages (O'Hara and Zhou, 2021). From a customer's perspective, the increasing electrification of the market implies that dealers can find a matching trading counterparty faster. Thus, I model this market innovation as a reduction in the execution delay of such mechanism [31]. Such delay is captured in the model by β. In the estimated calibration, customers wait on expectation one month to execute their trades. I use the model to analyze the impact of decreasing three times such delay. The new policy functions are presented in Figure 1.8.

Figure 1.8 shows that a reduction in the waiting for execution time affects the trading mechanism choice. Smaller execution delays imply that agency customers need to hold unwanted positions for less time, thus the relative attractiveness of such a contract increases. Consequently, customers with preference type - asset positions close to the baseline calibration thresholds migrate away from principal towards agency.

[31] Note that an alternative and non-mutually exclusive interpretation is a reduction in dealers' searching and matching costs, which are absent in my model.

Figure 1.8: Policy function as execution delays decrease



Note: This figure depicts the policy functions of each customer, conditional on her preference type and current holdings, considering $\beta = 3$. The lower and upper solid lines represent the buyer's and seller's optimal asset holdings under the principal trade, $a^{P,b}$ and $a^{P,s}$, respectively. The dashed line represents the optimal asset holdings under the agency trade, a^A. Regarding the mechanism choice, the principal and agency regions are shaded in orange and blue, respectively. To ease the comparison across calibrations, the trading mechanism thresholds under $\beta = 1$ are depicted as dotted lines within the agency region, and the arrows denote its expansion.

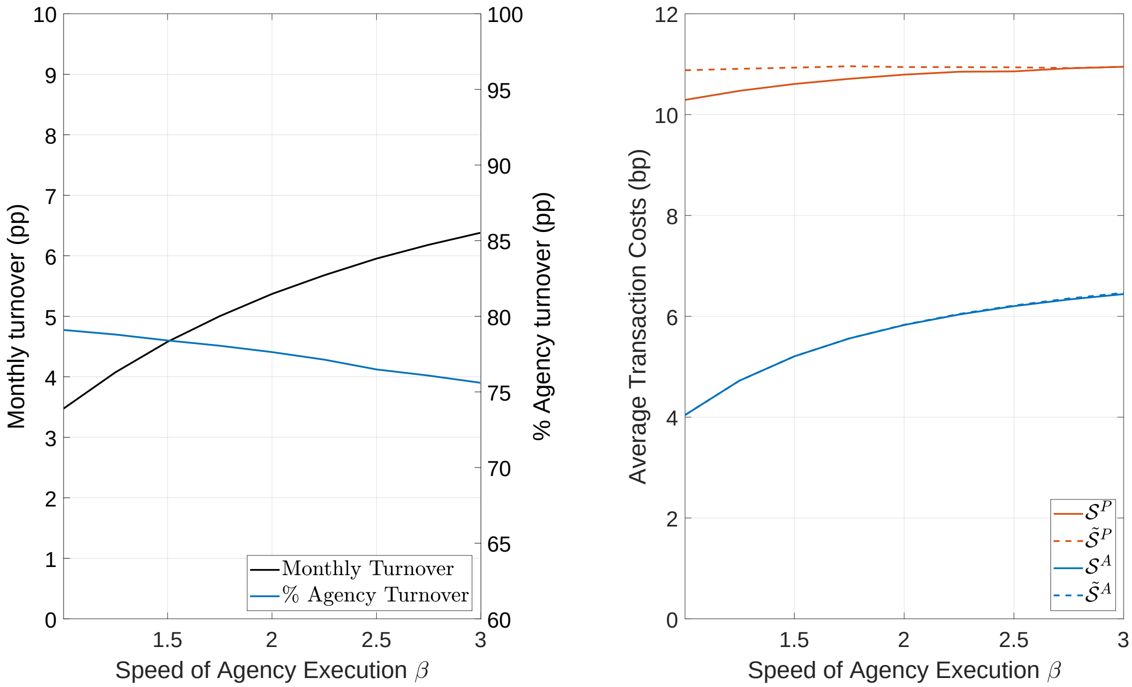
The liquidity measures computed for the range $\beta \in [1, 3]$ are presented in Figure 1.9. Panel A presents the daily turnover as well as the percentage explained by agency trades. Increasing the execution speed of non-immediate contracts largely affects the extensive margin of both principal and agency trading. On the one hand, the number of customers that signed an agency contract can trade faster. On the other hand, the mass of customers waiting for execution is reduced; therefore, more customers are able to contact dealers in any given month and optimally choose whether to arrange new principal or new agency contracts. A less obvious effect of reducing execution delays is the decrease in the intensive margin of agency trading compared to that of principal. Firstly, the migrating customers make the average volume traded in both principal and agency contracts larger. Figure 1.8 shows, for each preference type, the expansion of both the maximum and the minimum trading size under agency and principal trades, respectively[32]. Secondly, a faster execution implies that agency customers are more likely to avoid a preference shock while waiting for execution and trade according to their current preference types. Given that, in the steady state, the majority of the population is concentrated at the optimal asset positions, more customers trading according to the current type implies a decrease in the average agency volume per trade [33]. Overall, these effects jointly explain an increase in the daily turnover and a decrease in the agency share.

Panel B of Figure 1.9 shows the transaction costs in both mechanisms. Again, I decompose these figures into average and composition-free measures, which are depicted in solid and dashed lines, respectively. As execution delays decrease, average costs in both mechanisms go up. Principal costs increase by $\Delta\mathcal{S}^P = 0.66$ bps and agency costs rise by $\Delta\mathcal{S}^A = 2.40$ bps. Although speeding up agency trades makes trading in both mechanisms more expensive, the causes behind each of these changes are different. Regarding principal trades, the new calibration considered has no significant impact on the implied trading surplus of each customer. Therefore, keeping samples constant, principal costs should not significantly change. Accordingly, the counterfactual composition-free measure of principal costs has only a slight increase of $\Delta\tilde{\mathcal{S}}^P = 0.07$bp and almost the entire increase in average principal costs is due to composition effects, $CE^P = 89.54\%$. The explanation is found in Figure 1.8. Principal customers with relatively moderate preferences and asset positions, characteristics

[32]The optimal asset positions in the baseline and in the new calibration do not depart significantly, given that optimal assets are decided at execution in both mechanisms.

[33]LR09 contains a similar channel by which an increase in the contact rate with dealers, α, produces a steady state with a bigger accumulation of customers at their optimal positions, decreasing thus the average volume per trade.

Figure 1.9: Liquidity measures when execution delays decrease.



Note: Panel A (left) presents the steady-state total daily turnover rate, $\mathcal{T}$, and the agency percentage of such figure, $\mathcal{T}^{\mathcal{A}}/\mathcal{T}$, across $\beta \in [1, 3]$. Panel B (right) presents the steady-state volume-weighted average transaction costs for both mechanisms across $\beta \in [1, 3]$. Solid lines represent the average measures, $\mathcal{S}^{\mathcal{P}}$ and $\mathcal{S}^{\mathcal{A}}$, whereas dashed lines represent the counterfactual composition-free measures, $\tilde{\mathcal{S}}^{\mathcal{P}}$ and $\tilde{\mathcal{S}}^{\mathcal{A}}$.

associated with low transaction cost payments, migrate away from the mechanism, increasing the average willingness to pay of the remaining principal sample. Regarding agency trades, a reduction in expected delays has a direct positive impact on the expected trade surplus of every agency customer: unwanted positions can be exchanged faster. I compute an increase in the agency composition-free costs of $\Delta\tilde{\mathcal{S}}^A = 2.42$ bps. Note that this figure is slightly higher than the average measure, which indicates that inflowing migrating customers have a slightly smaller trade surplus than the non-migrant agency customers. The corresponding composition effect bias is negligible, computed at $CE^A = -1.03\%$.

The results here obtained provide new insights about the impact that electronic venues have in OTC markets. By reducing execution delays, these platforms produce a shift in the demand towards agency trades, thus raising the transaction costs of such a mechanism. An effect over principal costs is also observed, which operates exclusively through composition effects. As customers shift their demand towards agency, the sample of principal traders is reduced and the average surplus from trading on such mechanism increases. Therefore average immediacy costs spuriously increase. Although not studied here, the demand shifts observed arguably complement movements in the relative supply of trading mechanisms, due to the decrease in search and matching costs faced by dealers.

1.7 Conclusion

OTC markets have undergone several changes during the last decade. Intermediation activities had been perturbed by both new regulations and new trading technologies, affecting the cost and the speed at which customers can trade. In this paper, I study how customers optimally face these changing conditions and the consequences of such reaction over market liquidity and its measurement.

I develop a quantitative search model in which I can explicitly study the customers' trading mechanism choices. I show that the speed-cost trade-off faced when choosing between principal and agency trades is solved based on customers' trading needs, and that such trading needs are translated to transaction cost measures. The fact that trading mechanisms and transaction costs are jointly determined presents an empirical challenge. Whenever market conditions change, customers endogenously migrate across mechanisms, thus altering the composition of the samples in which liquidity measures are computed.

To overcome such challenge, I build counterfactual liquidity measures in which composition effects are controlled for. I estimate the model using corporate bond transaction data and perform numerical exercises motivated by recent developments in that market. In those exercises, a fraction of principal customers migrate towards agency trading. Given that those principal customers who did not migrate paid on average higher transaction costs, the change in principal average costs is upward biased. In particular, composition effects account for a third of the change in principal transaction costs after an inventory costs increase, and for almost all of the change after an increase in execution speed. In turn, agency costs are barely affected by composition effects.

The results here obtained contribute to the debate of whether stricter financial regulations set after 2008 were welfare-improving. If the cost of immediacy has not increased as much as was previously thought, new regulations may have improved financial soundness at a lower expense.

1.A Appendix: Model Additional Details

1.A.1 Bargaining Outcomes

Here I compute the bargaining outcomes for the principal contract. The agency contract terms of trade can be obtained similarly.

$$
\begin{aligned}
[a_i^P(a), \phi_i^P(a)] &= \underset{(a',\phi')}{\arg\max} \Big\{V_i(a') - V_i(a) - p(a'-a) - \phi'\Big\}^{1-\eta}\Big\{\phi' - \theta p|a'-a|\Big\}^{\eta} \\
&= \underset{(a',\phi')}{\arg\max}(1-\eta)\ln[\underbrace{V_i(a') - V_i(a) - p(a'-a) - \phi'}_{A}] + \eta \ln[\underbrace{\phi' - \theta p|a'-a|}_{B}].
\end{aligned}
$$

$$
\begin{aligned}
\text{FOC}_{\phi'}: \qquad & -(1-\eta)A^{-1} + \eta B^{-1} = 0 \qquad \text{(assume interior solution)} \\
& \eta A - (1-\eta)B = 0 \\
& \eta[V_i(a') - V_i(a) - p(a'-a)] + (1-\eta)\theta p|a'-a| = \phi_i^P(a)
\end{aligned}
$$

Second-order conditions can be checked trivially, therefore $\phi_i^P(a)$ is the unique global maximizer. Now let us introduce the solution for $\phi_i^P(a)$ in the maximization argument to obtain (1.4).

$$
\begin{aligned}
a_i^P(a) = \underset{a'}{\arg\max} & \Big\{(1-\eta)\big[V_i(a') - V_i(a) - p(a'-a) - \theta p|a'-a|\big]\Big\}^{1-\eta} \\
& \Big\{\eta\big[V_i(a') - V_i(a) - p(a'-a) - \theta p|a'-a|\big]\Big\}^{\eta} \\
\underset{a'}{\arg\max} \quad & V_i(a') - V_i(a) - p(a'-a) - \theta p|a'-a|.
\end{aligned}
$$

1.A.2 Customer's Value Function Using Bargain-adjusted Contact Rate.

Here I show that the customer's value function can be rewritten as if the contact rate with dealers was $(1-\eta)\alpha$ and the customer had full bargaining power. In other words, the utility flow of an investor trading at α rate with a dealer with η bargaining power is equal to that of an investor trading at a slower rate $(1-\eta)\alpha$ with a dealer with no bargaining power. Let me replace the optimal terms of trade from Equations (1.3), (1.4), (1.5) and (1.6) in Equation (1.1).

$$\begin{aligned}
V_{i(t)}(a) = \mathbb{E}_{i(t)}\Big[\int_t^{T_\alpha} & e^{-r[s-t]} u_{i(s)}(a) ds \\
& + e^{-r[T_\alpha - t]} \max\Big\{ (1-\eta)\big[V_{i(T_\alpha)}(a^P_{i(T_\alpha)}) - p(a^P_{i(T_\alpha)} - a) - \theta p |a^P_{i(T_\alpha)} - a|\big] + \eta V_{i(T_\alpha)}(a), \\
& \qquad (1-\eta) V^A_{i(T_\alpha)}\big(a, \phi^A_{i(T_\alpha)}(a) = 0\big) + \eta V_{i(T_\alpha)}(a) \Big\}\Big].
\end{aligned}$$

Define the time it takes for a customer to receive either the preference shock or the contact with dealers shock as τ_δ and τ_α, respectively. These are exponentially distributed with their corresponding parameters δ and α. In turn, define $\tau = \min\{\tau_\delta, \tau_\alpha\}$. Now consider the above Bellman equation over some small time horizon h, and let h go to zero:

$$\begin{aligned}
V_i(a) =& \frac{1}{1+rh}\Big[u_i(a)h + Pr[\tau = \tau_\alpha \le h]\Big[(1-\eta)\max\Big\{ V_i(a_i^P) - p(a_i^P - a) - \theta p |a_i^P - a|, \\
& \quad V_i^A\big(a, \phi_i^A(a) = 0\big)\Big\} + \eta V_i(a)\Big] + Pr[\tau = \tau_\delta \le h]\Big[\sum_j \pi_j V_j(a)\Big] + Pr[\tau > h] V_i(a)\Big] \\
=& \frac{1}{1+rh}\Big[u_i(a)h + \alpha h\Big[(1-\eta)\max\Big\{ V_i(a_i^P) - p(a_i^P - a) - \theta p |a_i^P - a|, V_i^A\big(a, \phi_i^A(a) = 0\big)\Big\} \\
& \quad + \eta V_i(a)\Big] + \delta h\Big[\sum_j \pi_j V_j(a)\Big] + (1 - \delta h - \alpha h) V_i(a)\Big]
\end{aligned}$$

$$=\frac{1}{1+rh}\Big[u_i(a)h + \underbrace{\alpha(1-\eta)h}_{Pr[\tau'=\tau_\kappa\leq h]}\left[\max\left\{V_i(a_i^P) - p(a_i^P - a) - \theta p|a_i^P - a|, V_i^A\big(a, \phi_i^A(a)=0\big)\right\}\right]$$

$$+\,\delta h\Big[\sum_j \pi_j V_j(a)\Big] + \underbrace{(1-\delta h - \alpha(1-\eta)h)}_{Pr[\tau'>h]} V_i(a)\Big],$$

where $\tau' = \min\{\tau_\delta, \tau_\kappa\}$ and τ_κ is the bargaining-adjusted time it takes to contact a dealer, which is exponentially distributed with parameter $\kappa = \alpha(1-\eta)$. Therefore, the customer's problem is represented by a Bellman equation where the contact with a dealer happens with Poisson arrival rate $(1-\eta)\alpha$, but where the customers have full negotiation power, $\eta' = 0$.

1.A.3 Expectations Resolution in the Flow Bellman Equation.

I keep on using τ_δ and τ_κ as the time it takes for a customer to receive either the preference shock or the (effective) contact shock, respectively, and $\tau' = \min\{\tau_\delta, \tau_\kappa\}$. In turn, define τ_β as the time it takes for a customer to be matched with another customer after choosing the agency trade. Consider the equation derived in Appendix 1.A.2 over some small time horizon h, and let h go to zero [34].

$$V_i(a) = \frac{1}{1+rh}\Big[u_i(a)h + Pr[\tau' = \tau_\delta \leq h]\sum_j \pi_j V_j(a)$$

$$+\,Pr[\tau' = \tau_\kappa \leq h]\max\left\{V_i(a_i^P) - p(a_i^P - a) - \theta p|a_i^P - a|, V_i^A(a)\right\} + Pr[\tau' > h]V_i(a)\Big]$$

$$V_i(a) = \frac{1}{1+rh}\Big[u_i(a)h + \delta h\sum_j \pi_j V_j(a)$$

$$+\,\kappa h\max\left\{V_i(a_i^P) - p(a_i^P - a) - \theta p|a_i^P - a|, V_i^A(a)\right\} + \big(1-(\delta+\kappa)h\big)V_i(a)\Big]$$

[34] For ease of exposition I removed time subscripts.

$$
\begin{aligned}
V_i(a)[\cancel{1} + r\cancel{h}] &= u_i(a)\cancel{h} + \delta\cancel{h}\sum_j \pi_j[V_j(a) - V_i(a)] \\
&\quad + \kappa\cancel{h}\max\Big\{V_i(a_i^P) - V_i(a) - p(a_i^P - a) - \theta p|a_i^P - a|, V_i^A(a) - V_i(a)\Big\} + \cancel{V_i(a)} \\
rV_i(a) &= u_i(a) + \delta\sum_j \pi_j[V_j(a) - V_i(a)] \\
&\quad + \kappa\max\Big\{V_i(a_i^P) - V_i(a) - p(a_i^P - a) - \theta p|a_i^P - a|, V_i^A(a) - V_i(a)\Big\},
\end{aligned}
$$

where $V_i^A(a)$ is the maximum utility a customer expects to get when she chooses the agency trade. Similarly, I can define this latter function in terms of flow utility as:

$$
rV_i^A(a) = u_i(a) + \delta\sum_j \pi_j[V_j^A(a) - V_i^A(a)] + \beta\big[V_i(a_i^A) - V_i^A(a) - p(a_i^A - a)\big],
$$

where $1/\beta$ is the time a customer expects to wait until the dealer finds him a counterpart and a_i^A is the optimal agency asset position chosen at execution (see Equation (1.6)). Note that, while waiting, the customer might change his preferences, which is reflected in the second term on the right-hand side of the above equation. The expression $V_i^A(a)$ can be further manipulated to be written as a function of $V_i(a)$. Let me first obtain the expression for $\sum_j \pi_j V_j^A(a)$:

$$
\begin{aligned}
(r + \delta + \beta)V_i^A(a) &= u_i(a) + \delta\sum_j \pi_j V_j^A(a) + \beta\big[V_i(a_i^A) - p(a_i^A - a)\big] \\
(r + \cancel{\delta} + \beta)\sum_i \pi_i V_i^A(a) &= \sum_i \pi_i u_i(a) + \delta\cancel{\sum_j \pi_j V_j^A(a)} + \beta\sum_i \pi_i\big[V_i(a_i^A) - p(a_i^A - a)\big] \\
\sum_j \pi_j V_j^A(a) &= \frac{1}{r + \beta}\Big[\sum_j \pi_j u_j(a) + \beta\sum_j \pi_j\big[V_j(a_j^A) - p(a_j^A - a)\big]\Big].
\end{aligned}
$$

Plugging this result into $V_i^A(a)$ equation:

$$(r+\delta+\beta)V_i^A(a) = u_i(a) + \frac{\delta}{r+\beta}\Big[\sum_j \pi_j u_j(a) + \beta\sum_j \pi_j\big[V_j(a_j^A) - p(a_j^A - a)\big]\Big] + \beta\big[V_i(a_i^A) - p(a_i^A - a)\big]$$

$$V_i^A(a) = \underbrace{\frac{1}{r+\beta}\frac{(r+\beta)u_i(a) + \delta\sum_j \pi_j u_j(a)}{r+\delta+\beta}}_{\bar{U}_i^\beta(a)} + \underbrace{\frac{\beta}{r+\beta}}_{\hat{\beta}}\Big[\underbrace{\frac{(r+\beta)V_i(a_i^A) + \delta\sum_j \pi_j V_j(a_j^A)}{r+\delta+\beta}}_{\bar{V}_i^A} - p\Big[\underbrace{\frac{(r+\beta)a_i^A + \delta\sum_j \pi_j a_j^A}{r+\delta+\beta}}_{\bar{a}_i^A} - a\Big]\Big]$$

$$V_i^A(a) = \bar{U}_i^\beta(a) + \hat{\beta}\big[\bar{V}_i^A - p(\bar{a}_i^A - a)\big]$$

Finally, I can include this result in the initial equation, rearrange and define terms in a similar way as was previously done. The flow Bellman equation of a customer of type i holding assets a waiting to contact a dealer in any given period is the following:

$$\begin{aligned} V_i(a) = {} & \\ & \bar{U}_i^\kappa(a) + \hat{\kappa}\Big[[1-\hat{\delta}]\max\Big\{V_i(a_i^P) - p(a_i^P - a) - \theta p|a_i^P - a|, \bar{U}_i^\beta(a) + \hat{\beta}\big[\bar{V}_i^A - p(\bar{a}_i^A - a)\big]\Big\} \\ & \quad + \hat{\delta}\sum_j \pi_j \max\Big\{V_j(a_j^P) - p(a_j^P - a) - \theta p|a_j^P - a|, \bar{U}_j^\beta(a) + \hat{\beta}\big[\bar{V}_j^A - p(\bar{a}_j^A - a)\big]\Big\}\Big], \end{aligned}$$

where $\bar{U}_i^\kappa(a) = \Big[\frac{(r+\kappa)u_i(a) + \delta\sum_j \pi_j u_j(a)}{r+\delta+\kappa}\Big]\frac{1}{r+\kappa}$, $\hat{\kappa} = \frac{\kappa}{r+\kappa}$ and $\hat{\delta} = \frac{\delta}{r+\delta+\kappa}$.

1.A.4 Trading Mechanism Choice Sets

After subtracting the common term $V_i(a)$, the indifference condition writes:

$$V_i(a_i^P) - p(a_i^P - a) - \theta p|a_i^P - a| = \bar{U}_i^{\beta}(a) + \hat{\beta}\big[\bar{V}_i^A - p(\bar{a}_i^A - a)\big]$$

Firstly, consider the indifference condition for the cases where agents change their positions should they trade under the principal mechanism. Conditional on increasing or reducing positions, and disregarding the current valuation $V_i(a)$, the gains from a principal trade increase at a constant rate in current asset holdings a. This is a direct consequence of modeling constant dealers' marginal costs and can be seen on the left-hand side of the indifference equation. On the other hand, in the agency mechanism, the customer keeps his current asset holdings until some counterparty is found. Given decreasing marginal instant utility, the flow utility she derives while waiting for execution, $\bar{U}_i^{\beta}(a)$, marginally decreases in current asset holdings a. After the waiting period is over, the customer will obtain a discounted gain from trade, which is also linear in a, since optimal agency holdings are independent of current holdings (see Equation (1.6)). Therefore, the total gains from a delayed intermediated trade are marginally decreasing in a. I will exploit these differences in the two types of trades to find the current asset holdings thresholds as the roots of the indifference condition. Let us rearrange the arguments of such an indifference equation:

$$\underbrace{V_i(a_i^P) - p(1+\psi\theta)a_i^P - \hat{\beta}(\bar{V}_i^A - p\bar{a}_i^A)}_{B_i} = \underbrace{\bar{U}_i^{\beta}(a)}_{C_i(a)} + \underbrace{pa\big(\hat{\beta} - (1+\psi\theta)\big)}_{D(a)},$$

where $\psi = 1$ $(= -1)$ if $a_i^P - a \geq 0$ (< 0). The left-hand side, B_i, is independent of current asset holdings a, while the two arguments on the right-hand side are not. Firstly, $C_i(a)$ is a twice continuously differentiable, strictly increasing, and strictly concave function that satisfies Inada conditions in current asset holdings a. Secondly, $D(a)$ is linear in a, and its sign depends on the difference between the expected present value of reselling the asset through agency and reselling the asset immediately plus the inventory cost discount. Given the assumption made about the marginal inventory costs, $\theta < \frac{r}{r+\beta}$, $D(a)$ is a strictly decreasing linear function on a, and the right-hand side is thus inverse U-shaped [35].

Let us consider now the indifference condition for the cases where customers would not trade if they were to opt for principal trading. A customer that does not trade derives utility by holding his current position until the next contact with a dealer. In turn, an agency trader adds up the utility of holding his current position until the execution of the trade, plus the gains from trade she gets without paying an immediacy premium. As before, I can rearrange this indifference condition to express its components according to their dependence on the current position.

$$\underbrace{-\hat{\beta}(\bar{V}_i^A - p\bar{a}_i^A)}_{B_i} = \underbrace{\bar{U}_i^{\beta}(a) - V_i(a)}_{C_i(a)} + \underbrace{pa\hat{\beta}}_{D(a)}.$$

The left-hand side, B_i, is still independent of current asset holdings a_i. Regarding the right-hand side, $D(a)$ is linear and strictly increasing in a. In turn, $C_i(a)$ subtracts from a strictly increasing and strictly concave function a function $V_i(a)$ that, at this point, is unknown. The shape of $C_i(a)$ determines the region under which customers decide not to trade at all. Given the unavailability of closed-form solutions for the value function, these regions are characterized numerically. Under all different plausible calibrations, the numerical solution of the model indicates that $C_i(a) + D_i(a)$ is U-shaped.

This analysis indicates that the optimal trading mechanism choice for each preference type can be characterized by partitions of the subsets $\Gamma_i = \{Buy_i, Sell_i, NoT_i\}$, which in turn

[35] The parameter values discussed in the calibration section indicate that $\theta < \frac{r}{r+\beta}$ is not a binding restriction for most plausible calibrations.

define the optimal trading direction for a customer trading on principal. Formally, define $\hat{a}_i^{h,\rho}$, with $h = \{1, 2\}$ and $\rho = \{b, s, nt\}$, as the current asset holdings that make customers of type i indifferent between the principal or the agency trade, where h denotes the threshold number and ρ indicates if the threshold is computed for a potential principal buyer, seller or non-trader. In turn, define the partitions Γ_i^P and Γ_i^A as the type-specific subsets of asset holdings within which a customer of type i would trade on principal or through agency in the steady state, respectively, for a specific principal trade direction $\Gamma_i = \{Buy_i, Sell_i, NoT_i\}$. The indifference condition provides two possible scenarios for each principal trade direction:

$$
Buy_i \begin{cases} B_i \geq C_i(a) + D_i(a) & \forall a: \qquad \Gamma_i^P = \Gamma_i. \\ B_i < C_i(a) + D_i(a) & \text{for some } a: \quad \Gamma_i^P = \Gamma_i \cap \{[-\infty, \hat{a}_i^{1,b}] \cup [\hat{a}_i^{2,b}, \infty)\}, \, \Gamma_i^A = \Gamma_i \setminus \Gamma_i^P. \end{cases}
$$

$$
Sell_i \begin{cases} B_i \geq C_i(a) + D_i(a) & \forall a: \qquad \Gamma_i^P = \Gamma_i. \\ B_i < C_i(a) + D_i(a) & \text{for some } a: \quad \Gamma_i^P = \Gamma_i \cap \{[-\infty, \hat{a}_i^{1,s}] \cup [\hat{a}_i^{2,s}, \infty)\}, \, \Gamma_i^A = \Gamma_i \setminus \Gamma_i^P. \end{cases}
$$

$$
NoT_i \begin{cases} B_i < C_i(a) + D_i(a) & \forall a: \qquad \Gamma_i^P = \emptyset. \\ B_i \geq C_i(a) + D_i(a) & \text{for some } a: \quad \Gamma_i^P = \Gamma_i \cap \{[\hat{a}_i^{1,nt}, \hat{a}_i^{2,nt}], \, \Gamma_i^A = \Gamma_i \setminus \Gamma_i^P. \end{cases}
$$

1.A.5 Existence and Uniqueness of the Value Function.

In order to prove the uniqueness of the value function $V_i(a)$, I need to show that the Bellman operator T, defined as the right-hand side of (1.7), is a contraction mapping that operates in a Banach space, i.e., a complete normed vector space. To show completeness, I can rely on Theorem 3.1 in Stokey, Lucas, and Prescott (1989) - SL89 -, which requires the functions mapped by T to be continuous and bounded. Define $S = R_+ \times \{1, .., I\}$,

$C = \{g : S \to R \mid g(a,i)$ is continuous in a and bounded above$\}$ and the metric space $(C, \|.\|)$, where $\|.\|$ denotes the *sup norm*. I want the right-hand side of Equation (1.7) to belong to C. By assumption, the utility function $u_i(a)$ is continuous, property preserved by the linear combination $\bar{U}_i^{\kappa}(a)$. Secondly, each term on the two sides of the max operator is continuous as well. Given the existence of thresholds $\bar{a}_i$ that make customers of type i indifferent between the two types of trade, both sides of the max operator return the same value at those thresholds. Hence, the utility a customer gets when her asset holdings change and cross a threshold does not suffer a jump. Finally, the stock of assets in the economy is in fixed supply $A \in R_+$, thus individual holdings are bounded. Therefore, $T : C \to C$ and $(C, \|.\|)$ is a complete metric space[36].

Our next step is to show that this operator is a contraction mapping. I will rely on Blackwell's sufficient conditions (Theorem 3.3, SL89). Therefore, I need to show that the operator satisfies the monotonicity and discounting properties.

Monotonicity: Take any pair $V^1, V^2 \in C$ such that $V^1(i,a) \leq V^2(i,a)$, for all {a,i}$\in S$. I need to show that $[TV^1](i,a) \leq [TV^2](i,a)$, for all {a,i}$\in S$. From Equation (1.7), the outcome of the max operators (decision of trade type) will always be greater or equal under $V^2(i,a)$ than under $V^1(i,a)$, since the arguments under both principal trade or agency are strictly increasing in the value function considered. The first term in Equation (1.7) does not depend on the value function, and the second term is a convex combination of these max operators (with weights $(1-\hat{\delta})$ and $\hat{\delta}$ respectively), so the weak inequality holds and monotonicity is achieved.

Discounting: I need to demonstrate that there exist some $\lambda \in (0,1)$ such that $[T(V +$

[36]The trading mechanism choice produces kinks in the value function. At those points, the value function will not be differentiable. Theorem 3.2 in SL89 only requires continuity, and that is guaranteed by the indifference condition that originates the kinks. See Kirkby (2017) for a proof of the convergence of the computational solution to the true solution using discretized value function iteration.

$\epsilon)](i,a) \leq [TV](i,a) + \lambda\epsilon$ for all $V \in C$, {a,i}$\in S$ and $\epsilon \geq 0$. Consider $[T(V+\epsilon)](i,a)$:

$$
\begin{aligned}
&[T(V+\boldsymbol{\epsilon})](i,a) = \\
&= \bar{U}_i^{\kappa}(a) \\
&+ \hat{\kappa}\Big[[1-\hat{\delta}]\max\Big\{V_i(a_i^P) - p(a_i^P - a) - \theta p|a_i^P - a| + \boldsymbol{\epsilon}, \bar{U}_i^{\beta}(a) + \hat{\beta}\big[\bar{V}_i^A - p(\bar{a}_i^A - a)\big] + \hat{\beta}\boldsymbol{\epsilon}\Big\} \\
&+ \hat{\delta}\sum_j \pi_j \max\Big\{V_j(a_j^P) - p(a_j^P - a) - \theta p|a_j^P - a| + \boldsymbol{\epsilon}, \bar{U}_j^{\beta}(a) + \hat{\beta}\big[\bar{V}_j^A - p(\bar{a}_j^A - a)\big] + \hat{\beta}\boldsymbol{\epsilon}\Big\}\Big] \\
&= \bar{U}_i^{\kappa}(a) \\
&+ \hat{\kappa}\Big[[1-\hat{\delta}]\max\Big\{V_i(a_i^P) - p(a_i^P - a) - \theta p|a_i^P - a|, \bar{U}_i^{\beta}(a) + \hat{\beta}\big[\bar{V}_i^A - p(\bar{a}_i^A - a)\big] - \mathbf{(1-\hat{\boldsymbol{\beta}})}\boldsymbol{\epsilon}\Big\} \\
&+ \hat{\delta}\sum_j \pi_j \max\Big\{V_j(a_j^P) - p(a_j^P - a) - \theta p|a_j^P - a|, \bar{U}_j^{\beta}(a) + \hat{\beta}\big[\bar{V}_j^A - p(\bar{a}_j^A - a)\big] - \mathbf{[1-\hat{\boldsymbol{\beta}}]}\boldsymbol{\epsilon}\Big\}\Big] \\
&+ \hat{\kappa}\boldsymbol{\epsilon} \\
&\leq [T(V)](i,a) + \hat{\kappa}\boldsymbol{\epsilon}
\end{aligned}
$$

where the last inequality comes from the fact that subtracting a scalar from a component of a max operator will yield a weakly smaller value. To gain intuition, consider the parametrization case such that all customers, i.e., any pair {a,i}, choose the principal trade. In that case, $[T(V+\epsilon)](i,a) \leq [TV](i,a) + \hat{\kappa}\epsilon$, where $\hat{\kappa} = \kappa/(r+\kappa) \in (0,1)$. Alternatively, consider the parametrization under which every customer chooses the agency trade. In such case, $[T(V+\epsilon)](i,a) \leq [TV](i,a) + \hat{\kappa}\hat{\beta}\epsilon$, where $\hat{\kappa}\hat{\beta} \in (0,1)$ as well. Any case in between will yield a discounting factor between these two bounds $[\hat{\kappa}\hat{\beta}, \hat{\kappa}]$.

1.A.6 Solution Method Algorithm

The steady state of the model for any given inter-dealer price, p, is solved using the value function iteration method, enhanced with Howard's improvement step. The obtained policy and value functions, conditional on p, are nested within the computation of the market clear condition 1.14 to obtain the equilibrium inter-dealer price. The algorithm can be described by the following steps:

1. Set an initial guess for the equilibrium price p.

    (a) Set an asset holdings grid and an initial guess for $V_i(a)$

    (b) Compute optimal asset holdings $\{a_i^P(a), a_i^A\}_{i=1}^I$ using Equations (1.4) and (1.6).

    (c) Compute trading mechanism choice for each pair $\{i, a\}$, using Equation (1.8).

    (d) Fix $\{a_i^P(a), a_i^A\}_{i=1}^I$, and iterate h times the following steps:

        i. Update $V_i(a)$ using Equation (1.7).

        ii. Compute trading mechanism choice for each pair $\{i, a\}$, using Equation (1.8)

    (e) Update $V_i(a)$ using Equation (1.7) until convergence.

2. Define trading mechanism sets $\{\Gamma_i^P, \Gamma_i^A\}_{i=1}^I$ using Equation (1.8).

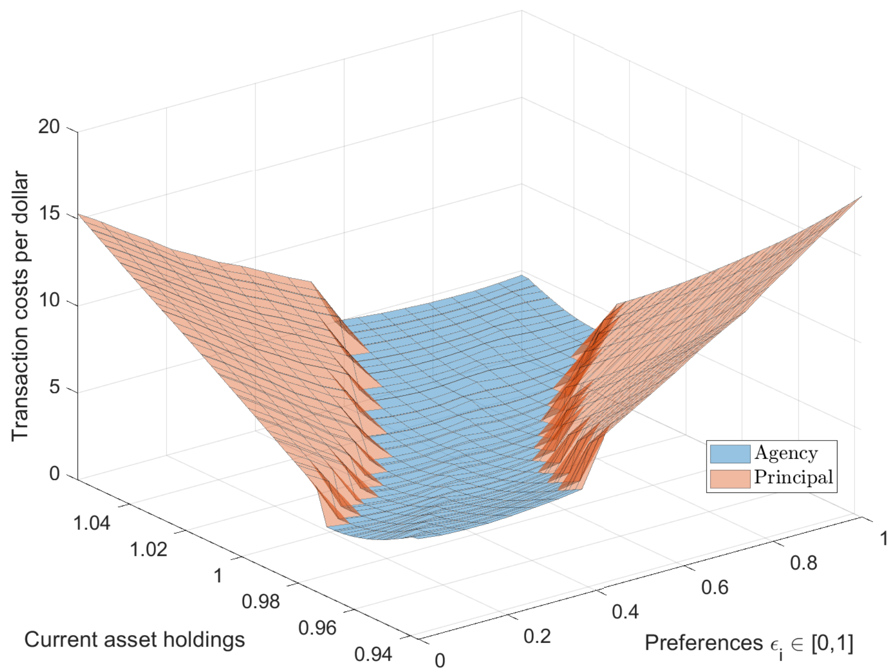
3. Compute transition matrix T using Equations (1.9), (1.10), (1.11), (1.12) and (1.13).

4. Set vector n_0 and obtain $n = \lim_{k \to K} n_0 T^k$, with K sufficiently large to reach convergence.

5. Compute aggregate gross demand and update p until excess demand in Equation (1.14) converges towards zero.

1.A.7 Transaction Costs per Dollar Traded

Figure 1.A.1: Transaction costs per dollar traded under each trading mechanism.



Note: This figure depicts the transaction costs per dollar traded paid by each customer, conditional on her preference type and current holdings, and expressed in basic points. Agency transaction costs are computed using the expected volume traded for each customer, as is explained in subsection 1.4.1, and expressed in present value at the moment of contact with the dealer.

1.A.8 Transaction Costs Decomposition

Here I present the algebra steps needed to decompose the transaction cost measures in Equations (1.17) and (1.18). Specifically, I decompose the transaction cost measures computed under some parametrization $q = 0$, considering an alternative parametrization $q = 1$. The decomposition of transaction costs computed for a different parametrization and considering a different alternative parametrization follow the same steps.

$$
\begin{aligned}
\mathcal{S}^{P,0} &= \sum_{i\in\mathcal{I}}\sum_{a\in P_i^0} \frac{n^0_{[a,i,\omega_1]}|a_i^{P,0}-a|}{\sum_{i\in\mathcal{I}}\sum_{a\in P_i^0} n^0_{[a,i,\omega_1]}|a_i^{P,0}-a|}\frac{\phi^{P,0}_{a,i}}{|a_i^{P,0}-a|p^0} \\
&= \underbrace{\sum_{i\in\mathcal{I}}\sum_{a\in P_i^0\cap P_i^1} \frac{n^0_{[a,i,\omega_1]}|a_i^{P,0}-a|}{\sum_{i\in\mathcal{I}}\sum_{a\in P_i^0\cap P_i^1} n^0_{[a,i,\omega_1]}|a_i^{P,0}-a|}\frac{\phi^{P,0}_{a,i}}{|a_i^{P,0}-a|p^0}}_{\mathcal{S}^{P,0}_{P^0,P^1}} \underbrace{\frac{\sum_{i\in\mathcal{I}}\sum_{a\in P_i^0\cap P_i^1} n^0_{[a,i,\omega_1]}|a_i^{P,0}-a|}{\sum_{i\in\mathcal{I}}\sum_{a\in P_i^0} n^0_{[a,i,\omega_1]}|a_i^{P,0}-a|}}_{w^{P,0}_{P^0,P^1}} \\
&+ \underbrace{\sum_{i\in\mathcal{I}}\sum_{a\in P_i^0\cap A_i^1} \frac{n^0_{[a,i,\omega_1]}|a_i^{P,0}-a|}{\sum_{i\in\mathcal{I}}\sum_{a\in P_i^0\cap A_i^1} n^0_{[a,i,\omega_1]}|a_i^{P}-a|}\frac{\phi^{P,0}_{a,i}}{|a_i^{P,0}-a|p^0}}_{\mathcal{S}^{P,0}_{P^0,A^1}} \underbrace{\frac{\sum_{i\in\mathcal{I}}\sum_{a\in P_i^0\cap A_i^1} n^0_{[a,i,\omega_1]}|a_i^{P,0}-a|}{\sum_{i\in\mathcal{I}}\sum_{a\in P_i^0} n^0_{[a,i,\omega_1]}|a_i^{P,0}-a|}}_{w^{P,0}_{P^0,A^1}} \\
&+ \underbrace{\sum_{i\in\mathcal{I}}\sum_{a\in P_i^0\cap NT_i^1} \frac{n^0_{[a,i,\omega_1]}|a_i^{P,0}-a|}{\sum_{i\in\mathcal{I}}\sum_{a\in P_i^0\cap NT_i^1} n^0_{[a,i,\omega_1]}|a_i^{P,0}-a|}\frac{\phi^{P,0}_{a,i}}{|a_i^{P,0}-a|p^0}}_{\mathcal{S}^{P,0}_{P^0,NT^1}} \underbrace{\frac{\sum_{i\in\mathcal{I}}\sum_{a\in P_i^0\cap NT_i^1} n^0_{[a,i,\omega_1]}|a_i^{P,0}-a|}{\sum_{i\in\mathcal{I}}\sum_{a\in P_i^0} n^0_{[a,i,\omega_1]}|a_i^{P,0}-a|}}_{w^{P,0}_{P^0,NT^1}} \\
&= \mathcal{S}^{P,0}_{P^0,P^1}w^{P,0}_{P^0,P^1} + \mathcal{S}^{P,0}_{P^0,A^1}w^{P,0}_{P^0,A^1} + \mathcal{S}^{P,0}_{P^0,NT^1}w^{P,0}_{P^0,NT^1}
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{S}^{A,0} &= \sum_{i\in\mathcal{I}}\sum_{a\in A_i^0} \frac{n^0_{[a,i,\omega_1]}rav^0_{a,i}}{\sum_{i\in\mathcal{I}}\sum_{a\in A_i^0} n^0_{[a,i,\omega_1]}rav^0_{a,i}}\frac{\phi^{A,0}_{a,i}}{rav^0_{[a,i]}p^0} \\
&= \underbrace{\sum_{i\in\mathcal{I}}\sum_{a\in A_i^0\cap A_i^1} \frac{n^0_{[a,i,\omega_1]}rav^0_{a,i}}{\sum_{i\in\mathcal{I}}\sum_{a\in A_i^0\cap A_i^1} n^0_{[a,i,\omega_1]}rav^0_{a,i}}\frac{\phi^{A,0}_{a,i}}{rav^0_{[a,i]}p^0}}_{\mathcal{S}^{A,0}_{A^0,A^1}} \underbrace{\frac{\sum_{i\in\mathcal{I}}\sum_{a\in A_i^0\cap A_i^1} n^0_{[a,i,\omega_1]}rav^0_{a,i}}{\sum_{i\in\mathcal{I}}\sum_{a\in A_i^0} n^0_{[a,i,\omega_1]}rav^0_{a,i}}}_{w^{A,0}_{A^0,A^1}} \\
&= \underbrace{\sum_{i\in\mathcal{I}}\sum_{a\in A_i^0\cap P_i^1} \frac{n^0_{[a,i,\omega_1]}rav^0_{a,i}}{\sum_{i\in\mathcal{I}}\sum_{a\in A_i^0\cap P_i^1} n^0_{[a,i,\omega_1]}rav^0_{a,i}}\frac{\phi^{A,0}_{a,i}}{rav^0_{[a,i]}p^0}}_{\mathcal{S}^{A,0}_{A^0,P^1}} \underbrace{\frac{\sum_{i\in\mathcal{I}}\sum_{a\in A_i^0\cap P_i^1} n^0_{[a,i,\omega_1]}rav^0_{a,i}}{\sum_{i\in\mathcal{I}}\sum_{a\in A_i^0} n^0_{[a,i,\omega_1]}rav^0_{a,i}}}_{w^{A,0}_{A^0,P^1}} \\
&= \underbrace{\sum_{i\in\mathcal{I}}\sum_{a\in A_i^0\cap NT_i^1} \frac{n^0_{[a,i,\omega_1]}rav^0_{a,i}}{\sum_{i\in\mathcal{I}}\sum_{a\in A_i^0\cap NT_i^1} n^0_{[a,i,\omega_1]}rav^0_{a,i}}\frac{\phi^{A,0}_{a,i}}{rav^0_{[a,i]}p^0}}_{\mathcal{S}^{A,0}_{A^0,NT^1}} \underbrace{\frac{\sum_{i\in\mathcal{I}}\sum_{a\in A_i^0\cap NT_i^1} n^0_{[a,i,\omega_1]}rav^0_{a,i}}{\sum_{i\in\mathcal{I}}\sum_{a\in A_i^0} n^0_{[a,i,\omega_1]}rav^0_{a,i}}}_{w^{A,0}_{A^0,NT^1}} \\
&= \mathcal{S}^{A,0}_{A^0,P^1}w^{A,0}_{A^0,P^1} + \mathcal{S}^{A,0}_{A^0,A^1}w^{A,0}_{A^0,A^1} + \mathcal{S}^{A,0}_{A^0,NT^1}w^{A,0}_{A^0,NT^1}
\end{aligned}
$$

where $rav^0_{a,i} = (1-\hat{\delta})|a_i^{A,0} - a| + \hat{\delta}\sum_{j\in\mathcal{I}} \pi_j |a_j^{A,0} - a|$.

1.B Appendix: Estimation Additional Details

1.B.1 Theoretical moments details

Here I describe how to compute the variances and covariances needed to calculate the slope between transaction costs and trade size. Let me start with the principal case.

$$cov\Big(\frac{10000\phi^P}{|a^P - a|p}, \frac{100|a^P - a|}{A}\Big) = \sum_{i\in\mathcal{I}}\sum_{a\in P_i} \frac{n_{[a,i,\omega_1]}}{\sum_{i\in\mathcal{I}}\sum_{a\in P_i} n_{[a,i,\omega_1]}} \Big(\frac{10000\phi^P_{a,i}}{|a_i^P - a|p} - \mathcal{S}^P_{nw}\Big)\Big(\frac{100|a_i^P - a|}{A} - \mathcal{V}^P\Big),$$
$$var(\frac{100|a_i^P - a|}{A}) = \sum_{i\in\mathcal{I}}\sum_{a\in P_i} \frac{n_{[a,i,\omega_1]}}{\sum_{i\in\mathcal{I}}\sum_{a\in P_i} n_{[a,i,\omega_1]}} \Big(\frac{100|a_i^P - a|}{A} - \mathcal{V}^P\Big)^2$$

where $\mathcal{S}^P_{nw}$ is the non-weighted average principal transaction costs and $\mathcal{V}^P$ is the average principal trade size:

$$\mathcal{S}^P_{nw} = \sum_{i\in\mathcal{I}}\sum_{a\in P_i} \frac{n_{[a,i,\omega_1]}}{\sum_{i\in\mathcal{I}}\sum_{a\in P_i} n_{[a,i,\omega_1]}} \Big(\frac{10000\phi^P_{a,i}}{|a_i^P - a|p}\Big)$$
$$\mathcal{V}^P = \sum_{i\in\mathcal{I}}\sum_{a\in P_i} \frac{n_{[a,i,\omega_1]}}{\sum_{i\in\mathcal{I}}\sum_{a\in P_i} n_{[a,i,\omega_1]}} \Big(\frac{100|a_i^P - a|}{A}\Big)$$

For the case of agency trades:

$$cov\Big(\frac{10000\phi^A}{rav\times p}, \frac{100rav}{A}\Big) = \sum_{i\in\mathcal{I}}\sum_{a\in A_i} \frac{n_{[a,i,\omega_1]}raf_{a,i}}{\sum_{i\in\mathcal{I}}\sum_{a\in A_i} n_{[a,i,\omega_1]}raf_{a,i}} \Big(\frac{10000\phi^A_{a,i}}{rav_{a,i}\times p} - \mathcal{S}^A_{nw}\Big)\Big(\frac{100rav_{a,i}}{raf_{a,i}}\frac{1}{A} - \mathcal{V}^A\Big),$$
$$var(\frac{100rav}{A}) = \sum_{i\in\mathcal{I}}\sum_{a\in A_i} \frac{n_{[a,i,\omega_1]}raf_{a,i}}{\sum_{i\in\mathcal{I}}\sum_{a\in A_i} n_{[a,i,\omega_1]}raf_{a,i}} \Big(\frac{100rav_{a,i}}{raf_{a,i}}\frac{1}{A} - \mathcal{V}^A\Big)^2$$

where $\mathcal{S}^A_{nw}$ is the non-weighted average agency transaction costs, $\mathcal{V}^A$ is the average agency trade size, and $raf_{a,i}$ is the realized agency fraction of customers in state $n_{[a,i,\omega_1]}$ who actually end up trading, i.e., those who hold asset holdings different than their optimal at execution:

$$\mathcal{S}^A_{nw} = \sum_{i\in\mathcal{I}}\sum_{a\in A_i} \frac{n_{[a,i,\omega_1]} raf_{a,i}}{\sum_{i\in\mathcal{I}}\sum_{a\in A_i} n_{[a,i,\omega_1]} raf_{a,i}} \left(\frac{10000\phi^A_{a,i}}{rav_{a,i}\times p}\right)$$
$$\mathcal{V}^A = \sum_{i\in\mathcal{I}}\sum_{a\in A_i} \frac{n_{[a,i,\omega_1]} raf_{a,i}}{\sum_{i\in\mathcal{I}}\sum_{a\in A_i} n_{[a,i,\omega_1]} raf_{a,i}} \left(\frac{100 rav_{a,i}}{raf_{a,i}}\frac{1}{A}\right)$$
$$raf_{a,i} = (1-\hat{\delta})\mathbf{1}_{a^A_i\neq a} + \hat{\delta}\sum_{j\in\mathcal{I}} \pi_j \mathbf{1}_{a^A_j\neq a}.$$

1.B.2 Regression

Here I present the estimation results for the equation

$$s_{t,d,b} = \alpha + \beta FE + \gamma 100(vol_{t,b,d}/iao_b) + \epsilon_{t,b,d},$$

where $s_{t,b,d}$ is Choi, Huh, and Seunghun Shin (2024)'s Spread1, $vol_{t,b,d}$ is the volume traded, iao_b is the bonds' average amount outstanding, and $FE = [dealer, bond, day]$. The data employed as well as the principal/agency distinction is explained in subsection 1.5.2.

Table 1.B.1: transaction costs - trade size regressions.

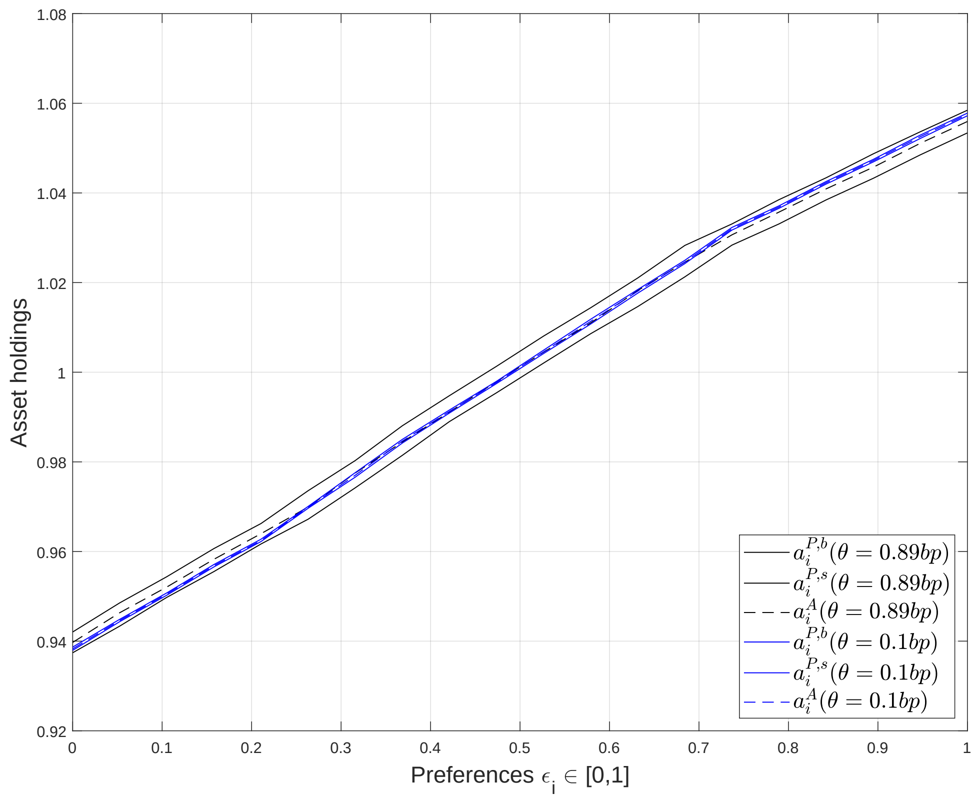
	Principal	Agency
Trade size (pp)	1.45***	0.61***
	(0.13)	(0.12)
Dealer FE	Yes	Yes
Bond FE	Yes	Yes
Day FE	Yes	Yes
Observations	1,505,133	97,305
R^2	0.111	0.019

Note: This table provides OLS estimates of the trade-level regression of Choi, Huh, and Seunghun Shin (2024)'s measure of transaction costs Spread1 on $100vol/iao$ ratio, dealer fixed effects, bond fixed effects and day fixed effects, where $vol_{t,b,d}$ is the volume traded and iao_b is the bonds' average amount outstanding. The model is estimated for principal and agency trades separately. Clustered day-bond standard errors are shown in parentheses. One, two, and three stars indicate statistical significance at the 0.1, 0.05, and 0.01, respectively.

1.C Appendix: Numerical Exercises Additional Details

1.C.1 Optimal Assets with Low and High Inventory Costs

Figure 1.C.1: Optimal asset as inventory costs increase.



Note: This figure depicts the optimal asset positions of each customer, conditional on her preference type and current holdings, considering $\theta = 0.1$ bps and $\theta = 0.89$ bps. The lower and upper solid lines represent the buyer's and seller's optimal asset holdings under the principal trade, $a^{P,b}$ and $a^{P,s}$, respectively. The dashed line represents the optimal asset holdings under the agency trade, a^A. The cases for low and high inventory costs are in blue and black, respectively.

1.C.2 Quantitative Exercises Robustness Checks

This appendix presents the composition effects (CE) computed for both quantitative exercises, using alternative values of externally calibrated parameters. I consider alternative preference distributions, with $\pi_i \sim Beta(\lambda, \lambda)$, and alternative dealer's bargaining power η. The parameters not affected are kept at their baseline calibration value.

Table 1.C.1: Composition Effects under alternative calibrations

		Composition Effect					
		λ			η		
		0.2	1	5	0.91	0.95	0.99
$\Delta\theta$	CE^P	18.49	32.19	28.65	25.99	32.19	34.58
	CE^A	-0.20	-1.19	0.42	0.50	-1.19	-16.78
$\Delta\beta$	CE^P	79.64	89.54	101.38	74.71	89.54	105.18
	CE^A	-1.14	-1.03	0.26	-1.09	-1.03	-4.08

Note: This table presents the composition effects resulting from increasing inventory costs θ from 0.1bps to 0.89bps (rows 1 and 2) and from increasing the agency rate β from 1 to 3 (rows 3 and 4), computed for alternative preference distributions, using $\pi_i \sim Beta(\lambda, \lambda)$, and alternative dealer's bargaining power η. The parameters not affected are kept at their baseline calibration values.

Chapter 2

Portfolio Trading in OTC Markets: Transaction Cost Discounts and Penalties

This paper studies a recent innovation in the corporate bond market: portfolio trading. In contrast to sequential trading, this new protocol allows customers to trade a list of bonds as a single security. I show that these trading features have significant consequences on market liquidity. Particularly, I present novel evidence of asymmetrical transaction costs: compared to sequential trading, portfolio trading is less expensive when customers buy bonds and more expensive when they sell them. I find that dealers' balance sheet costs and portfolios' diversification explain such differences.

2.1 Introduction

The corporate bond market has undergone several transformations in recent years. Market participants have shifted from trading through voice messages to doing so on electronic platforms (Hendershott and Madhavan, 2015; O'Hara and Zhou, 2021), dealers have accommodated stricter regulations by relying more on pre-arranged trades instead of trading with their inventories (Bessembinder, Jacobsen, Maxwell, and Venkataraman, 2018; Choi, Huh, and Seunghun Shin, 2024), and all-to-all platforms, where customers can skip dealer intermediation, are becoming increasingly popular (Hendershott, Livdan, and Schürhoff, 2021). The latest of these innovations is portfolio trading, a new protocol in which market participants can bundle a set of bonds and trade them as a single security. Although involving higher commitment from dealers, the electronic platforms that supply this new protocol claim that portfolio trading helps not only to improve execution quality but also to reduce transaction costs.[1]

In this paper, I study portfolio trading in the corporate bond market, addressing to what extent the claims made by electronic platforms hold empirically. I start by addressing the evolution of portfolio trading. I develop an algorithm to infer portfolios from individually reported trades and find that, starting in 2018, this new protocol has become increasingly popular, both in the customer-dealer and in the inter-dealer segment. Its provision is highly concentrated among top dealers, who rely on inventories to provide liquidity. I next turn to study the cost of portfolio trading. Compared to traditional sequential trading, customers sell portfolios with a penalty and buy portfolios at a discount. These transaction cost differences are explained by two forces: the overall volume and the overall risk traded. Moreover, portfolio penalties and discounts are not distributed homogeneously across bonds. I find a significant cross-subsidy within portfolios, where the traditional bond characteristic pricing is reversed once a bond is included in a portfolio.

The first task I perform is to infer portfolio trades from the Trade Reporting and Compliance Engine (TRACE). This database only recently (May 2023) adopted a protocol identifier, thus I need to develop an algorithm to track portfolios back in time. In a nutshell,

[1]See, for example, providers Tradeweb and ICE portfolio trading descriptions.

I look for two counterparts executing many different bonds in the same second. As expected, the algorithm captures the rise of portfolio trading in early 2018, the period when platforms started offering the service. I find that portfolio trading accounts for more than 10 billion dollars of monthly volume during late 2020, evenly divided between the customer-dealer and the inter-dealer segments, capturing 5% of the total market. Recent estimates show that the positive trend continued during 2021 (Li, O'Hara, Rapp, and Zhou, 2023).

To understand this new protocol, I provide descriptive statistics comparing portfolio and sequential trading. Portfolios are mainly institutional trades, typically involving around one hundred bonds and 65 million dollars of face value. Not surprisingly, its intermediation is concentrated among top dealers, which have enough sophistication to price all bonds and inventories to back up these trades. I find that portfolios affect dealers' balance sheets, whether because bonds were held in inventories before selling them to customers or because bonds add up to inventories after being bought in portfolios. Regarding portfolio composition, I do not find evidence suggesting that customers use this protocol to sell low-turnover bonds, as portfolios have a lower concentration of low-turnover bonds and a higher concentration of mid-turnover bonds than sequential trades. I do find evidence of customers trading riskier bonds in portfolios, measured both by interest rate risk and credit risk.

I next turn to study whether portfolio trading improves or hinders liquidity. To guide the empirical analysis, I provide a theoretical framework of transaction costs in over-the-counter (OTC) markets. Trading bonds through portfolios instead of sequentially would increase (decrease) transaction costs if it increases (decreases) customers' trading surpluses or dealers' trading costs. Several portfolio characteristics that may drive these variables are outlined. Among them, a portfolio implied balance sheet cost, how much risk a portfolio can diversify, and the likelihood of customers trading on private information at least one of the bonds included in the portfolio.

The empirical analysis starts by comparing the transaction costs paid by customers when trading bonds sequentially or through portfolios, controlling for other relevant characteristics of the trade. I find that bonds traded in portfolios pay on average 17.7% less transaction costs than those traded sequentially. However, the effect is asymmetric. When customers buy portfolios from dealers, they have a 42.6% transaction cost discount. Contrastingly, when customers sell portfolios to dealers, they pay a 9.9% penalty. These results hold robustly when considering alternative model specifications and alternative sample periods.

To understand what factors are behind these discounts and penalties, I proceed in two ways. On the one hand, I address how individual bonds are priced within the portfolios. I find a significant cross-subsidy within portfolios: characteristics that are priced in sequential trading are reversed when bonds are included in a portfolio. On the other hand, I investigate what portfolio characteristics are priced by dealers and in which direction. I find significant evidence of both balance sheet effects and portfolio diversification effects. Bonds in large-size portfolios have associated transaction costs up to 36.34 basis points (bps) higher than those in small-size portfolios. In turn, bonds in portfolios with many bonds – proxy for risk diversification – pay up to 27.67 bps less to trade than bonds in portfolios with few lines.

Overall, portfolio trading appears as a disruptive innovation in the corporate bond market. It provides a better execution quality for those customers in need of trading many bonds simultaneously. However, such improvement in execution quality does not always come for free. As this paper shows, when customers sell portfolios to dealers, they incur an extra cost compared to that of trading bonds sequentially. These higher costs can be further exacerbated if portfolios involve large volumes and do not diversify individual bond risk.

2.1.1 *Related Literature*

This paper is related to two strands of the literature. On the one hand, it complements the empirical literature that studies recent developments in the corporate bond market. For example, the rise of electronic platforms (Hendershott and Madhavan, 2015; O'Hara and Zhou, 2021) and all-to-all trading (Hendershott, Livdan, and Schürhoff, 2021), the effect of stricter banking regulations after the global financial crisis (Anderson and Stulz, 2017; Bao, O'Hara, and Zhou, 2018; Bessembinder, Jacobsen, Maxwell, and Venkataraman, 2018; Dick-Nielsen and Rossi, 2019; Choi, Huh, and Seunghun Shin, 2024; Rapp and Waibel, 2023), or episodes of big turmoil as COVID-19 (Kargar, Lester, Lindsay, Liu, Weill, and Zúñiga, 2021). I address how the latest innovation in this market, portfolio trading, is used by customers and dealers and how it affects market liquidity. On the other hand, this paper informs the theoretical literature on OTC markets, which models trading in a sequential fashion (Duffie, Gârleanu, and Pedersen, 2005; Lagos and Rocheteau, 2009; Weill, 2020), and the long-standing theoretical literature on portfolio pricing (Markowits, 1952; Acharya and Pedersen, 2005), which assumes assets can be traded continuously. In this regard, portfolio trading offers a unique opportunity to study the pricing of OTC-traded portfolios. I show that portfolio trading is associated with higher costs when customers sell and lower costs when customers buy, and provide the factors behind these asymmetries. Finally, this paper closely relates to two independent, contemporaneous works on corporate bonds portfolio trading. Meli and Todorova (2022) use proprietary data to study investment-grade portfolios. They find that transaction costs are reduced by over 40% when trading portfolios. I complement their findings by incorporating the whole universe of portfolios, both investment-grade and high-yield, and showing that portfolio transaction costs can be larger than sequential costs, especially for large-size and less diversified portfolios. In turn, Li, O'Hara, Rapp, and Zhou (2023) use the regulatory version of TRACE and find that portfolios are usually traded at a discount, although that discount is reduced the more balance sheet dealers accumulate as a result of the portfolio buy. My results show that customers pay higher costs when selling portfolios than when doing it so sequentially, and that the size of a portfolio increases its costs, both when customers buy and when they sell, indicating that dealers also translate the balance sheet costs of those bonds that were held in inventory before the trade.

2.2 Portfolio Trading: a New Protocol in the Bonds Market

The US corporate bond market is a typical over-the-counter market, where the lack of a centralized exchange makes customers search for trading counterparties. Typically, dealers reduce these search frictions by intermediating transactions, using their own inventories and locating counterparts within their trading network. Although communications have shifted from phone calls and Bloomberg messages, i.e. voice trading, to electronic platforms, customer-dealer interactions can still be described in the same following steps. Customers would contact dealers requesting quotes, specifying the issue, the trade size, and whether is a buy or a sell order. Dealers with the capability of providing quotes would compete, and the best quote would execute the trade. Since neither receiving quotes from dealers nor executing the trade at the winning quote is guaranteed, the execution uncertainty adds up to the search friction as a major concern for customers in this market.

In many scenarios, customers would like to trade many bonds simultaneously, e.g. portfolio rebalancing, fixed income exchange-traded funds (ETF) create and redeem process, etcetera. In such cases, customers would need to contact dealers sequentially, repeating the process previously described for each bond. In practice, customers engage in list trading: they send a spreadsheet with all the orders to dealers, who choose whether to offer quotes or not on a bond-by-bond basis. As these quotes are usually not firm, the process often suffers many back-and-forth iterations until all bonds are traded, turning list trading into a long and laborious practice.

As an improving alternative to sequential trading, electronic platforms such as ICE, MarketAxess, and Tradeweb started offering a new trading protocol called portfolio trading. This protocol allows customers to bundle a list of bonds and trade them as a single security. Through electronic platforms, customers put dealers in competition requesting quotes for the entire portfolio of bonds, in an all-or-none fashion. If the customer agrees, the portfolio is executed at the best quote received.

Compared to sequential trading, portfolio trading offers a better execution quality, as it reduces the time it takes to execute all desired trades and guarantees that all bonds within the portfolio are executed. Notwithstanding these benefits, electronic platforms claim that portfolio trading also minimizes information leakage, as the number of dealers contacted to execute all bonds would be reduced, and helps to trade illiquid bonds, which dealers would not be willing to trade unless structured into a bigger package. Moreover, portfolio trading is supposed to be cheaper than sequential trading. The argument behind such a claim is that, through portfolio diversification, customers reduce the risk dealers are asked to trade, and so the pricing of the bonds included in the portfolio improves. In the following sections I test many of these claims.

2.3 Data and Portfolio Trading Summary Statistics

In this section I describe the data used and how I identify portfolio trades. I show that portfolio trades represent a significant and growing fraction of the market and that its intermediation is concentrated among top dealers, who source bonds using their balance sheets. Finally, I provide relevant summary statistics comparing portfolio and sequential trades.

2.3.1 Data

I rely on three databases to study portfolio trading in the corporate bond market. The first and main data source is the academic version of the TRACE database, produced by the Financial Industry Regulatory Authority (FINRA). This data contains all corporate bond secondary market transactions reported by broker-dealers registered as member firms of FINRA. Importantly, the academic version of TRACE contains dealers' identifiers, which allows me to infer portfolio trades out of bundled trades. I extend this data with the Mergent Fixed Income Securities Database (FISD), which contains a broad set of bond characteristics not present in TRACE. Finally, I obtain complementary time-series variables from the Federal Reserve Economic Data (FRED). The period considered spans from January 2016 to December 2019.

To produce the final data set, I start by filtering TRACE out of reporting errors, duplicated observations, and book-keeping observations. This database is built out of reported trades, and thus it may contain reporting errors. I follow the procedure outlined in Dick-Nielsen and Poulsen (2019) to remove such errors [2]. I further remove duplicated inter-dealer trades, i.e. trades that are reported twice as both counterparts are reporting dealers. Finally, I delete those trades in which dealers transfer bonds to their non-FINRA affiliates for book-keeping purposes (Adrian, Boyarchenko, and Shachar, 2017).[3]

Next, I extend the filtered database by adding bond-level variables from FISD. To remove idiosyncratic features of bond contracts that may bias the transaction costs analysis, I follow the empirical literature and apply several filters (e.g., Bessembinder, Jacobsen, Maxwell, and Venkataraman, 2018; Friewald and Nagler, 2019; Kargar, Lester, Lindsay, Liu, Weill, and Zúñiga, 2021). Among them, the most significant ones are dropping bonds that

[2] Both the algorithm and the filter results can be downloaded from my personal website.
[3] Starting on November 2, 2015, FINRA provides explicit labels for the so-called book-keeping trades.

are preferred, convertible or exchangeable, yankee bonds, bonds with a sinking fund provision, variable coupon, with time to maturity of less than a year, or issued less than two months before the transaction date. I also remove bonds that are security-backed, equity-linked, putable, denominated in foreign currency, privately placed, perpetual, sold as part of a unit deal, or secured lease obligations bonds.

Finally, aiming at capturing only institutional investors, I remove trades of less than ten thousand dollars in face valuation. In this regard, Pinter, Wang, and Zou (2024) shows that transaction costs paid by retail and institutional investors significantly differ. As it will be shown in subsection 2.3.3, unlike sequential trades, portfolio trades are mostly institutional-size trades. Therefore, removing small trades allows for a fair comparison between sequential and portfolio trades. The final database is composed of 24,782,434 observations from 15,231 different bonds.

2.3.2 Portfolio Trades Identification

The structure of the data requires a strategy to identify portfolio trades. On the one hand, every bond traded is reported as a single observation, regardless of the trading protocol used, i.e. portfolio or sequential trading. On the other hand, there is no trading protocol flag for the period analyzed in this study. In this regard, although electronic platforms started offering portfolio trading in early 2018, its potential economic significance among scholars and researchers has been acknowledged only recently. As a result, an explicit flag for portfolio trades is absent in TRACE for observations reported before May 15, 2023.[4] In the following paragraphs, I describe how I identify portfolio trades by using observations' characteristics.

A portfolio trade is the exchange of a bundle of bonds by two counterparts at a unique

[4]See FINRA Regulatory Notice 22-12.

price. Clearly, the characteristics of this trade impose several restrictions on the individual reporting of the bonds that form the portfolio. I use these restrictions to identify portfolio trades. First, all bonds should be traded at the same time. Second, only two counterparts should be involved in the transaction. Third, the amount of bonds traded should be enough to be considered a bundle. Finally, there should be no duplicated bonds within a portfolio. The following algorithm identifies as portfolio trades those bundles of individual reports that satisfy the aforementioned restrictions:

1. Build bundles of bonds traded by the same dealer, in the same second, against the same counterpart.
2. Remove the duplicated bonds within those bundles.
    a If the counterpart is another dealer:
        i. Remove all duplicated bonds
    b If the counterpart is a customer:
        i. Remove those duplicated bonds that have the same trade side, i.e. buy or sell.
        ii. Keep bundles in which there are no duplicated bonds or where all bonds are duplicated with observations of the same volume but with opposite trade sides.
3. Tag as portfolio trades those bundles that, after the duplicated bonds removal, include ≥ 30 bonds.

In the first step of the algorithm, I build bundles of bonds that are traded by the same dealer, in the same second, against the same counterpart. In the second step, I clean those bundles from duplicated bonds. As can be seen, this latter step treats inter-dealer and customer-dealer trades differently. This responds to the lack of customer identifiers in TRACE. Specifically, when the algorithm requires to build bundles of bonds traded against the same counterpart, in the case of customer-dealer trades we cannot be sure if the bonds are being traded with a unique customer or with many customers. Thus, I cannot remove all duplicated bonds in a customer-dealer bundle, as I may be dealing with a case where a dealer buys a portfolio from a customer and sells the same portfolio to another customer at the same time. Instead, I approach the removal of duplicated bonds in customer-dealer bundles in two steps. First, I remove only those duplicated bonds that have the same trade side, i.e. buy or sell. This step mostly captures bundles formed entirely by observations of the same bond and same trade side (95.46% of observations removed belong to such bundles). Second, after the removal of duplicated bond-side observations, I only keep bundles with no duplicated bonds or those in which we can clearly observe two symmetric buy and sell portfolios. Finally, the third step is a portfolio minimum-size filter consistent with the discussions held by the Securities Industry and Financial Markets Association (SIFMA) and FINRA about the appropriate threshold to trigger a portfolio trading flag in TRACE. [5]

The strategy to identify portfolio trades is in line with strategies used by other authors. Meli and Todorova (2022) matches proprietary data on investment-grade portfolio requests for quotes with TRACE. With those matched observations, they build a clustering algorithm that resembles the one here presented. In turn, Li, O'Hara, Rapp, and Zhou (2023) uses TRACE and improves over the clustering algorithm of Meli and Todorova (2022) performing different refinements. Among these refinements, their algorithm deletes all duplicated bonds in a cluster, thus mechanically removing any customer-dealer portfolio trade that is immediately offloaded with another customer. By capturing those portfolios, I can speak to the sourcing of portfolios and how they impact transaction costs.

[5]See SIFMA response to FINRA's Regulatory Notice 20-24 - Proposed Changes to TRACE Reporting Relating to Delayed Treasury Spot and Portfolio Trades.

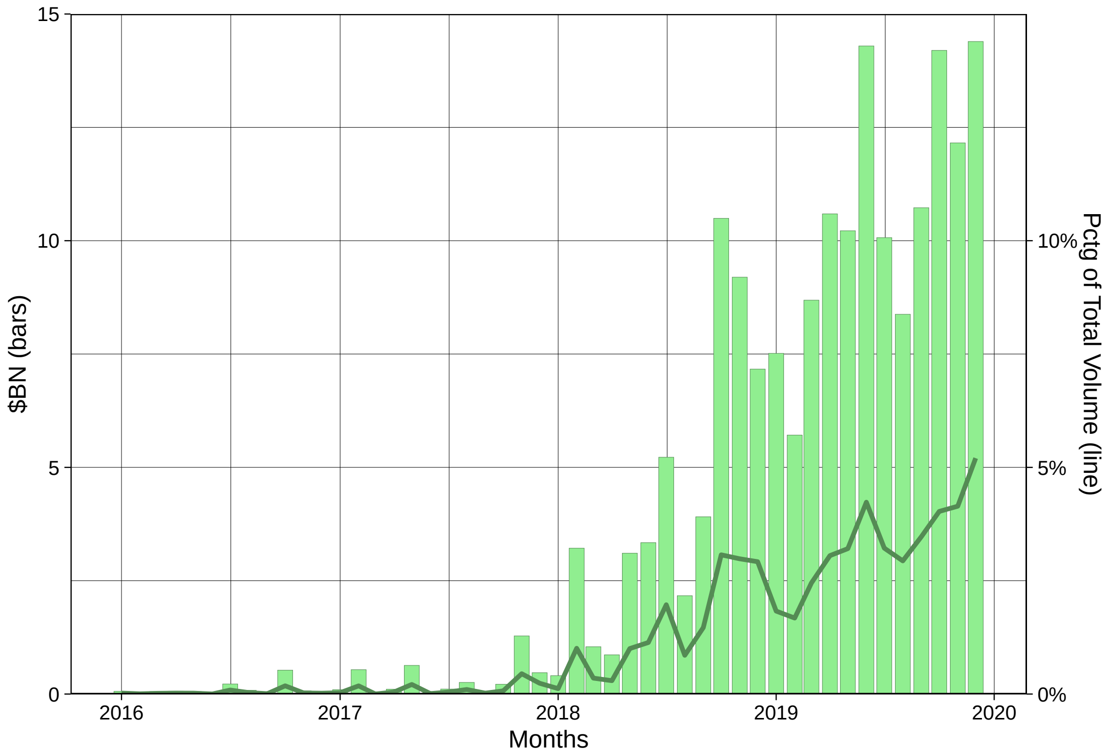
Finally, it is worth noting that the algorithm to identify portfolio trades better suits an environment of infrequent trading. In other words, if dealers execute several transactions every second, a bundle of sequential trades randomly executed at the same second against the same counterpart could be mistakenly inferred as a portfolio trade. This is especially problematic in the case of customer-dealer inferred portfolios, as the counterpart identity is unknown. In the Appendix 2.A.1 I show that dealers do not trade frequently. Particularly, Table 2.A.1 shows, for those dealers that perform portfolio trades, both how many seconds pass by between two customer-dealer trades and how many customer-dealer trades are performed in every second in which at least one trade is performed. The distribution of these variables shows that bundles of more than 30 bonds traded between dealers and customers are a rare event, which only happens at the extreme tail of the distribution.

Figure 2.1 shows the monthly evolution of the identified portfolio trading volume. As expected, the identified portfolio trading volume sharply rose in early 2018, i.e. when electronic platforms started offering the protocol, reaching more than 10 billion dollars of monthly trading during the second half of 2019. The market share mimics this pattern, reaching 5% of the total volume traded in the secondary corporate bond market. In the Appendix 2.A.2 I show that these patterns hold in the two market segments, i.e. inter-dealer and customer-dealer, and if we consider the number of trades instead of volume.

2.3.3 Portfolio Characteristics

In this subsection, I present descriptive statistics of the portfolios identified in subsection 2.3.2. I restrict the sample in two ways. On the one hand, since the main focus of this paper is to study transaction costs, I restrict the analysis to customer-dealer trades. Two reasons explain this decision. First, in the customer-dealer segment it is clear who demands liquidity (customers) and who provides it (dealers). Thus, transaction costs reflect the price paid to

Figure 2.1: Portfolio trading volume - All segments.



Note: This figure depicts the monthly time-series of portfolio trading volume, including both customer-dealer and inter-dealer trades. The bars –left axis– indicate total face value, expressed in billion dollars. The line –right axis– indicates market share, expressed in percentage points.

dealers to supply liquidity. Second, I will show that the trade side is a leading factor of transaction costs, and this variable is only relevant when we know who is providing liquidity. On the other hand, as electronic platforms started offering the portfolio trading alternative in early 2018, I restrict the sample to the period that goes from January 2018 to December 2019. The final sample consists of 7,633,744 customer-dealer individual trades, including 1,558 portfolios that account for 154,587 of those trades.

I start by addressing the size of portfolios. Table 2.1 shows that these are typically comprised of around one hundred bonds, although they can reach up to more than three hundred issues. The bonds in a portfolio are usually distributed across several issuers.

Regarding the volume traded, it is clear that portfolio trading is performed by institutional investors: the average portfolio involves 65.4 million dollars and 95% of portfolios involve more than 2 million dollars (face value).

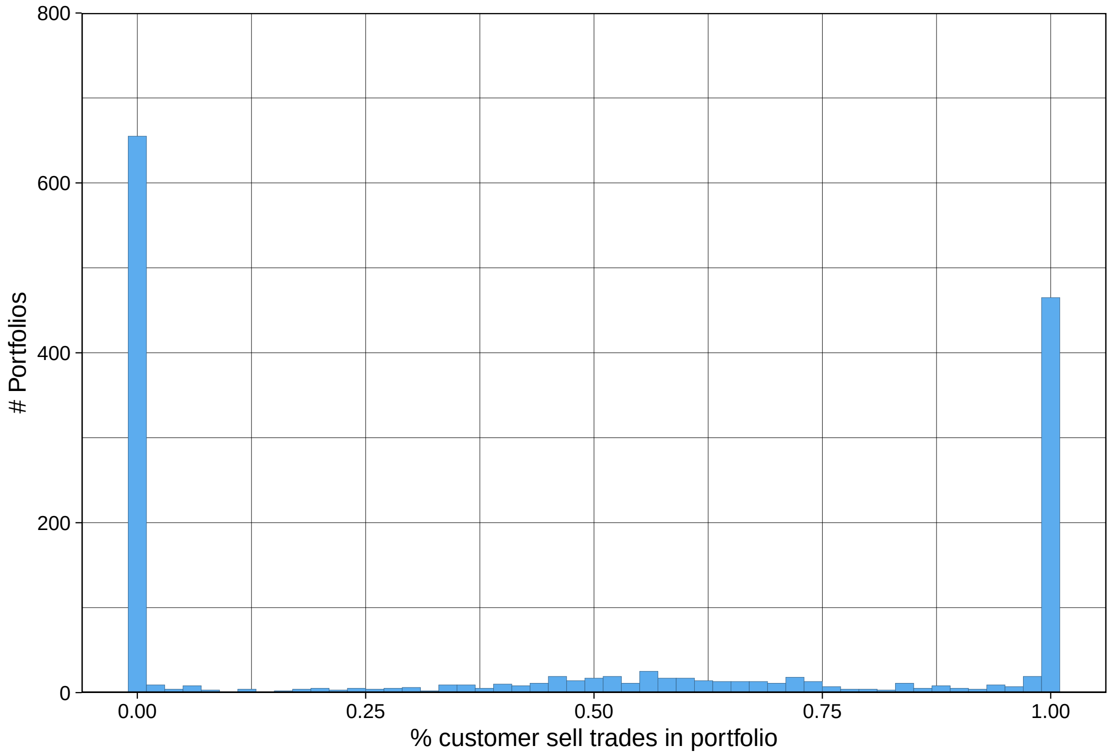
Table 2.1: Portfolios Size.

	Mean	Std. dev.	.05	.25	.50	.75	.95
Bonds #	99.2	108.5	31.0	39.0	57.0	109.0	323.2
Issuers #	74.5	65.0	27.0	35.0	49.0	86.0	213.2
Portfolio Size $M	65.4	155.3	2.1	8.9	23.0	58.8	255.4
Trades Size $M	0.81	1.75	0.04	0.14	0.34	0.68	3.00

Next, I turn to the question of whether customers use portfolio trading to buy, sell, or change the composition of the bonds they hold. This is particularly relevant as electronic platforms allow to mix buy and sell orders within a portfolio, thus customers can use the protocol to rebalance positions avoiding a timing mismatch between buying and selling and the risk implied by it. Figure 2.2 shows that portfolio trading is used for different strategies. 42% of portfolios are full customer buys and 30% are full customer sells, representing 40% and 28% of the portfolio volume of our sample. The remaining fraction is composed of mixed portfolios.

A small caveat should be mentioned at this point. Given my portfolio identification strategy, if a dealer decides to upload the buy orders and the sell orders of a mixed portfolio at different times, I will consider that mixed portfolio as two independent buy and sell portfolios. Although rare, Meli and Todorova (2022), by matching portfolio requests for quotes with actual trades from TRACE, shows that such cases exist. To address this concern, I combine those buy and sell portfolios executed by the same dealer within a 15-minute window (the maximum time allowed by FINRA to report trades after execution), and obtain that only 4% of the full buy or full sell portfolios can be considered as two legs of mixed portfolios. Turning to the supply side, I observe that portfolio trading is highly concentrated among top

Figure 2.2: Share of customer sell trades in portfolios.



Note: This figure depicts the distribution of customer sell trades percentage in portfolios. For each portfolio, I compute the percentage of customer sell trades. Bars express the number of portfolios with a certain percentage of customer sells.

dealers. The top three portfolio dealers accumulate 86% of the volume traded (87.46% of the bonds traded). These dealers happen to account for a large market share in the sequential protocol as well, suggesting that only big sophisticated dealers are able to price and trade the large number of bonds and volume implied by portfolios. In Appendix 2.A.3 I show that this concentration is stable over time, although the market shares of some dealers fluctuate, as is expected with any new technology.

The aforementioned market concentration is related to how bonds are sourced. In this regard, I find that dealers use their balance sheets when performing portfolio trades. To get this result, I follow the literature (Bessembinder, Jacobsen, Maxwell, and Venkataraman,

Table 2.2: Concentration of dealer intermediation of portfolio trades.

Dealer	Trades % share Portfolio	Trades % share Sequential	Volume % share Portfolio	Volume % share Sequential
1	35.29	6.66	49.93	10.20
2	21.17	4.29	18.65	8.88
3	31.00	1.73	17.40	0.71
4	3.17	3.12	6.50	8.09
5	2.32	2.86	3.33	8.51
6	3.29	3.95	2.46	7.49
7	1.46	2.72	0.72	8.18
8	0.27	1.76	0.25	5.35
9	0.30	0.09	0.20	0.03
10	0.21	0.32	0.16	0.34

2018; Kargar, Lester, Lindsay, Liu, Weill, and Zúñiga, 2021; Choi, Huh, and Seunghun Shin, 2024) and classify all customer-dealer trades into three categories: those that are quickly offset with other customers, those that are quickly offset with other dealers, and those that are not offset. Specifically, for each customer-dealer trade, I look for all the offsetting trades of the same dealer in the same bond, within a 15-minute window. If at least 50% of its volume was offset, and the majority of such volume was offset with customers (dealers), I label it as "Offset $\leq$ 15 - C" ('Offset $\leq$ 15 - D"). If less than 50% of its volume was offset, I label it as "Non-Offset". [6] Only this last "Non-Offset" category of trades affects dealers' balance sheets. Table 2.3 shows that the large majority of bonds traded through portfolios belong to such a category. These figures are much higher than those of sequential trading, where dealers tend to offset a larger fraction with other customers. These results are in line with the high concentration of portfolio trading among large dealers, as these are the ones with large enough balance sheet capacity to accommodate portfolios. [7]

[6] Two subtleties about this categorization are worth mentioning. First, this procedure allows for multiple matching, in the sense that a single trade can be offset by several trades of the opposite direction. Second, the algorithm may encounter competing trades. In such case, I form pairs with the trades that are closer in time firstly, and closer in volume secondly.

[7] These patterns hold if we perform the categorization using a 30-minute window, or if we consider the number of trades instead of the volume traded. See Appendix 2.A.4.

Table 2.3: Sourcing of Portfolio - Volume.

	Market Share		Portfolio Sourcing			Sequential Sourcing		
			Offset ≤15m		Non-Offset	Offset ≤15m		Non-Offset
Dealer	Portfolio	Sequential	C	D		C	D	
1	49.9	10.2	3.5	1.7	94.8	16.6	3.4	80.0
2	18.6	8.9	4.3	1.7	94.0	17.4	3.3	79.3
3	17.4	0.7	0.0	1.7	98.3	0.0	1.6	98.4
4	6.5	8.1	3.2	1.2	95.6	16.6	3.4	79.9
5	3.3	8.5	14.7	0.3	85.0	21.7	2.3	76.0
6	2.5	7.5	0.7	0.3	99.0	15.9	2.5	81.5
7	0.7	8.2	0.1	1.6	98.4	18.1	2.5	79.4
8	0.2	5.3	0.0	0.1	99.9	24.1	3.7	72.2
9	0.2	0.0	0.0	99.5	0.5	0.6	70.4	29.0
10	0.2	0.3	0.0	100.0	0.0	0.0	100.0	0.0

Note: This table shows, for each of the top ten portfolio trading dealers, its portfolio trading market share (column 2), its sequential trading market share (column 3), the distribution in the three categories – Offset ≤ 15 - C, Offset ≤ 15 - D, Non-Offset – of its portfolio trading activity (columns 4-6) and sequential trading activity (columns 7-9). All statistics are computed using volume traded, measured at face value.

Finally, I turn to the characteristics of the bonds included in a portfolio. In Table 2.4, I look at trade size, turnover, time to maturity, and credit rating, comparing how these variables are distributed in the portfolio and sequential trading subsamples. Appendix 2.A.5 explains in detail the construction of these variables. As previously noted, portfolio trading is mostly formed by institutional-size trades. Whereas more than 60% of sequential trades do not surpass 100 thousand dollars, less than 30% of portfolio trades belong to that category. Surprisingly, portfolios do not seem to be biased towards bonds with smaller turnover. Electronic platforms claim that portfolio trading could improve the liquidity of low-turnover bonds, as packaging helps dealers mitigate the risk of miss-pricing bonds for which transactions are rare.[8] I cannot find evidence supporting such a claim. Finally, we observe that portfolios have a somewhat higher concentration of riskier bonds, both considering time to maturity as a proxy for interest rate fluctuation risk and (to a lesser extent) credit risk.

[8] See for example electronic platform Tradeweb's "Portfolio Trading: An Innovative Solution for Corporate Bond Trading".

This higher concentration of riskier bonds in portfolios is not surprising, as the implied diversification reduces the overall risk of the position.

Table 2.4: Trade characteristics of portfolio and sequential trades.

	Trades % share		Volume % share	
	Portfolio	Sequential	Portfolio	Sequential
Trade Size				
Micro (≤100K)	29.24	60.76	2.52	2.63
Odd (100K-1M)	59.02	24.32	34.44	12.60
Round (1M-5M)	10.19	11.75	34.61	40.49
5M and above	1.55	3.17	28.44	44.28
Turnover				
(0%-10%]	18.30	25.47	21.34	17.43
(10%-25%]	43.29	45.46	44.55	40.17
(25%-50%]	29.28	20.52	26.49	27.70
>50%	9.13	8.54	7.63	14.70
Time to Maturity				
(1-3]	9.96	22.47	11.57	15.70
(3-5]	21.32	23.17	20.50	19.47
(5-10]	46.36	37.24	43.14	40.50
>10	22.36	17.12	24.78	24.33
Rating				
IG superior	3.98	6.52	5.44	5.30
IG inferior	42.50	65.99	57.35	59.10
HY superior	48.59	23.51	33.96	28.06
HY inferior	4.93	3.98	3.25	7.54

Note: This table shows how portfolio trades and sequential trades are distributed across partitions of trade size, turnover, time to maturity, and credit rating. The first two columns compute percentages using the number of trades. The last two columns compute percentages using the face value volume traded.

2.4 Transaction Costs

By construction, portfolio trading offers some advantages to those customers seeking to trade many bonds. For example, the protocol binds customers from holding temporary unwanted positions that would occur if they were to trade the bonds sequentially. Notwithstanding, it has been argued that portfolio trading is also cheaper than sequential trading, as dealers provide better prices for portfolios than for the sum of the individual bonds that compose them. In this section, I study such a claim. I start by providing a theoretical framework of transaction costs. Later, I provide trade-level evidence of transaction cost differences between portfolio and sequential trading and which factors drive those differences.

2.4.1 Theoretical Framework

To study whether portfolios are traded at a discount or penalty, I start by providing a theoretical framework that explains how transaction costs are settled. I follow the bulk of the literature on OTC markets and assume that the terms of trade are the outcome of bilateral bargaining, a natural assumption as counterparts in this market trade bilaterally instead of in a centralized exchange. [9] In particular, I assume that the quantity traded q and the transaction cost that a customer pays to a dealer $\phi(q)$ are solved through Nash bargain[10]:

$$[q^*, \phi(q)^*] = \arg\max_{(q,\phi)} \Big\{\mathrm{CS}(q) - \phi(q)\Big\}^{1-\eta} \Big\{\phi(q) - \mathrm{DC}(q)\Big\}^{\eta}$$

where CS and DC denote the customer surplus and the dealer cost, respectively, and $\eta \in [0, 1]$ reflects the dealer's bargaining power. The solution to this maximization problem tells us that, if there are gains from trade (CS>DC), the resulting transaction cost is a convex combination of the dealer cost and the customer surplus:

$$\phi(q)^* = \eta\mathrm{CS}(q) + (1 - \eta)\mathrm{DC}(q) \tag{2.1}$$

[9] For a review of this literature, see Weill (2020)

[10] Duffie, Gârleanu, and Pedersen (2007) model explicitly a bilateral bargaining game where agents make alternate offers. They show that the powers of the Nash product equal the probabilities of making an offer in such a game.

Equation (2.1) reveals that the effect of portfolio trading over transaction costs will be explained by how this new protocol affects customers' surpluses and dealers' costs. On the one hand, the customer surplus is increased due to better execution quality. As previously mentioned, when customers need to trade many bonds, they may be temporarily exposed to unwanted positions while all their trades are executed. Portfolio trading allows for simultaneous execution, thus avoiding such a risk. This larger consumer surplus should translate into higher transaction costs for portfolios. On the other hand, the dealers face different costs when trading portfolios or trading sequentially. First, as it was documented in subsection 2.3.3, portfolios are large institutional-size trades that affect dealers' balance sheets. In contrast with sequential trading, where a dealer can gradually offset positions keeping its balance sheet close to its target, portfolio trading implies large deviations from it. These deviations are costly to dealers (e.g. regulatory cost) and thus should translate into higher transaction costs. Second, portfolios comprise a large number of bonds issued by several firms. The resulting diversification of expected payoffs reduces the amount of risk being traded, decreasing thus dealers' costs. The more diversified a portfolio is, the smaller the transaction costs we should expect to observe. Last but not least, portfolio trading may be used by customers who have private information about some assets but do not want to signal it through an individual order. Dealers may anticipate this strategy and penalize the entire portfolio. Consequently, this channel would decrease the transaction costs of the bonds for which private information is held and increase that of the remaining bonds.

In the next subsection, I initially answer whether portfolios are trading at a penalty or at a discount. Later, I study the drivers behind the differences found, following the hypotheses aforementioned.

2.4.2 Transaction Costs Discounts and Penalties

Transaction costs are computed as the Spread1 measure of Choi, Huh, and Seunghun Shin (2024). Particularly, the transaction cost TC compares each customer-dealer trade price with a reference price, the latter given by the (volume-weighted) average price that the same bond has during the same day in the inter-dealer market.[11]

$$TC_{i,b,d} = Q \times (\frac{p_{i,b,d} - p^{DD}_{b,d}}{p^{DD}_{b,d}}) \times 10,000 \qquad , \qquad p^{DD}_{b,d} = \frac{\sum_{i \in DD_{b,d}} vol^{DD}_{b,d,i} p^{DD}_{b,d,i}}{\sum_{i \in DD_{b,d}} vol^{DD}_{b,d,i}}$$

where i, b, and d denote trade, bond, and day, respectively, Q is a trade side indicator that equals 1 (-1) if the customer buys (sells) bonds, and the multiplication by 10.000 expresses transaction costs as basis points deviations from the inter-dealer price.[12]

As a first approximation, in Table 2.5 I present how transaction costs are distributed within the portfolio and sequential subsamples. Clearly, those bonds that are traded within portfolios do so at smaller transaction costs: while the average transaction cost in portfolio trading is 8.6 bps, the average cost in sequential trading is 31.3 bps.

Table 2.5: Transaction costs by trade type.

	Transaction Costs (bps)						
	Mean	Std. dev.	.05	.25	.50	.75	.95
Portfolio	8.6	41.5	-42.5	-7.4	5.9	23.0	67.7
Sequential	31.3	82.3	-19.3	0.5	10.9	37.7	164.4

[11] Alternative transaction costs measures had been used in the empirical fixed income literature, among them Amihud (2002) price impact and Feldhütter (2012) round trip costs. The accuracy of these measures relies on having close-in-time consecutive trades of the same bond, a feature hardly observed in the portfolio trading subsample.

[12] As is the case with any measure of transaction costs, the elements needed for its construction restrict the sample for which we can compute it. In this case, the only restriction is for dealer-customer trades to match with an inter-dealer trade of the same bond happening on the same day. In Appendix 2.A.6 I show how such restriction affects the samples of portfolio trades and sequential trades.

Of course, these transaction cost differences may be driven by factors other than the inclusion of a trade in a portfolio. I improve the analysis by computing the transaction costs differential associated with the inclusion of a bond in a portfolio trade, conditional on several bond and trade characteristics. Specifically, I estimate through OLS the following empirical model:

$$TC_i = \alpha + \beta \mathbf{1}_{i=\text{Portfolio}} + \Gamma C_i + \Lambda FE + \epsilon_i, \tag{2.2}$$

where TC_i denotes the transaction cost of trade i, the dummy variable $\mathbf{1}_{i=\text{Portfolio}}$ indicates if such trade belongs to a portfolio trade, and the vectors C_i and FE includes bond and trade characteristics and several fixed effects, respectively. Regarding bond and trade characteristics, I control for age, amount outstanding, time to maturity (TTM), credit rating, trade size, and whether the trade was performed by a dealer who performs portfolio trading.[13]. In turn, the model includes day, issuer industry, dealer, and bonds fixed effects, which are used according to each specification of Equation (2.2). Standard errors are double clustered by bond and date.

The first column of Table 2.6 presents the baseline estimation results. The coefficient associated with including a bond in a portfolio, controlling for several priced characteristics, is negative and significant. The transaction cost of a bond executed through portfolio trading is expected to be 5.53 bps smaller than that of a bond executed through sequential trading. Taking into account the mean transaction costs presented in Table 2.5, this represents a 17.7% discount. The results show that transaction costs are also led by the type of dealer that intermediates: dealers who trade portfolios (typically big dealers) charge smaller transaction costs. The coefficients associated with the remaining controls are in line with previous findings in the literature (e.g., Edwards, Harris, and Piwowar, 2007). Bonds issued in large

[13]See Appendix 2.A.5 for the detailed computation of these variables.

Table 2.6: Transaction costs regression on trade characteristics.

Dependent Variable:	Transaction Cost					
	Baseline	Dealer FE	Bond FE	No DST	No Offset	No Mixed
Portfolio	-5.53***	-4.54***	-3.32***	-4.91***	-6.24***	-4.87***
	(0.74)	(0.64)	(0.67)	(0.84)	(0.76)	(0.89)
Portfolio Dealer	-26.11***		-20.75***	-26.08***	-26.63***	-26.07***
	(0.45)		(0.36)	(0.45)	(0.47)	(0.45)
Age	0.04	-0.26***		0.03	0.02	0.03
	(0.13)	(0.09)		(0.13)	(0.13)	(0.13)
Amount Outs.	-2.71***	-1.93***		-2.71***	-2.75***	-2.71***
	(0.40)	(0.31)		(0.40)	(0.40)	(0.40)
TTM 3-5	7.76***	6.97***		7.75***	8.05***	7.77***
	(0.60)	(0.51)		(0.60)	(0.61)	(0.60)
TTM 5-10	18.74***	14.98***		18.77***	19.41***	18.78***
	(0.72)	(0.55)		(0.72)	(0.73)	(0.72)
TTM >10	48.49***	36.06***		48.65***	49.76***	48.63***
	(1.53)	(0.96)		(1.53)	(1.55)	(1.53)
Odd (100K-1M)	-20.23***	-8.51***	-16.68***	-20.29***	-20.13***	-20.33***
	(0.43)	(0.22)	(0.37)	(0.43)	(0.43)	(0.43)
Round (1M-5M)	-28.32***	-12.82***	-22.82***	-28.47***	-28.10***	-28.49***
	(0.65)	(0.39)	(0.56)	(0.65)	(0.64)	(0.65)
5M and above	-23.79***	-9.44***	-19.19***	-23.99***	-22.65***	-23.98***
	(0.62)	(0.37)	(0.49)	(0.62)	(0.63)	(0.62)
IG (A-BBB)	8.04***	3.63***		8.06***	8.20***	8.06***
	(0.67)	(0.46)		(0.67)	(0.68)	(0.67)
HY (BB-B)	24.17***	15.73***		24.25***	25.34***	24.27***
	(0.99)	(0.67)		(0.99)	(1.02)	(0.99)
HY (CCC-D)	48.34***	39.85***		48.48***	52.17***	48.54***
	(3.33)	(3.06)		(3.33)	(3.57)	(3.34)
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Issuer Industry FE	Yes	Yes	No	Yes	Yes	Yes
Dealer FE	No	Yes	No	No	No	No
Bond FE	No	No	Yes	No	No	No
Observations	6,300,985	6,300,985	6,307,999	6,279,622	6,021,275	6,276,809
Adjusted R^2	0.095	0.202	0.141	0.095	0.109	0.095
Within R^2	0.093	0.030	0.031	0.093	0.107	0.092

Note: This table provides OLS estimates of the trade-level Equation (2.2). The baseline specification regresses transaction cost on a portfolio trade dummy, a portfolio dealer dummy, age, amount outstanding, time to maturity, credit rating, trade size, day fixed effects and issuer industry fixed effects. Alternative specifications include dealer fixed effects (column 2), bond fixed effects (column 3), the exclusion of portfolios executed within a 5-minute window of delayed spot times (column 4), the exclusion of "Offset$\leq 15m$ - C" trades (column 5), and the exclusion of mixed portfolios (column 6). Clustered day-bond standard errors are shown in parentheses. One, two, and three stars indicate statistical significance at the 0.1, 0.05, and 0.01, respectively.

amounts have smaller transaction costs, as these are easier to price and trade. Bonds far away from maturity are more expensive to trade, a result related to these bonds having higher interest rate risk and more uncertainty in their valuation. Larger trades (> 100K) are cheaper than smaller trades, driven probably by the identity of the investors, a variable not available in my data sets (Pinter, Wang, and Zou, 2024). As expected, worse credit-rated bonds are traded at higher transaction costs, as dealers translate the implied risk cost to customers.

The main result of transaction costs being smaller for those trades included in portfolio trading holds under alternative specifications. I firstly account for dealer-heterogeneity and its effect on transaction costs (e.g. Colliard, Foucault, and Hoffmann, 2021). The second column of Table 2.6 shows that the result holds when imposing dealers' fixed effects. To fully account for bond time-insensitive characteristics, in specification three I include bond fixed effects. I observe that the portfolio discount holds, although to a lesser extent than in the baseline model. I also consider a specification where I remove those portfolio trades executed within a 5-minute window of popular delayed spot times: 11.00, 15.00, 15.30, 16.00, and 16.30. These times of the day are used to execute trades that had been priced as a spread over some reference price, leading thus to an accumulation of trades that may be mistakenly inferred as portfolio trading. Column four tells us that removing those observations does not affect the results. Another robustness check performed is to remove from the sample trades that are offset within a 15-minute window with other customers. In this kind of trade, there are no dealers' balance sheets involved, and thus transaction costs are typically smaller. As such trades are more prevalent in sequential trading, its presence in the sample would underestimate the portfolio trading discount. The estimated coefficient of column five confirms the claim. In addition to the previous robustness checks, I estimate the model using only full buy or full sell portfolios. Mixed portfolios may not imply balance sheet cost, as buy and sell orders net out, removing one of the channels that affect transaction costs.

Again, column six shows that results hold robustly. Finally, given that portfolio trading is a new protocol, it may be the case that dealers initially offered better pricing as a strategy to gain market power. In that case, the discount observed would not be sustained when the market matures. In untabulated estimations, I see that all results hold if we restrict the sample to the period June 2019 to December 2019, discarding thus this hypothesis.

Considering that, among the four hypotheses cited, only portfolio diversification would reduce transaction costs, it is surprising to see a discount holding robustly across all specifications. To further understand this result, I study whether customers pay different transaction costs when buying or selling portfolios. If dealers' balance sheet costs respond asymmetrically to deviations from the target, e.g. penalizing more positive deviations than negative ones, it would be expected to observe a portfolio trading asymmetric effect on transaction costs. I formally test for asymmetric effects by estimating an extended version of Equation (2.2):

$$TC_i = \alpha + \beta_1 \mathbf{1}_{i=\text{Portfolio}} + \beta_2 \mathbf{1}_{i=\text{Cust. sells}} + \beta_3 \mathbf{1}_{i=\text{Portfolio}} \mathbf{1}_{i=\text{Cust. sells}} + \Gamma C_i + \Lambda FE + \epsilon_i, \quad (2.3)$$

Equation (2.3) decomposes the portfolio trading subsample into those trades in which customers buy and those in which customers sell bonds, with associated coefficients β_1 and $\beta_1 + \beta_3$, respectively. The estimation results are presented in Table 2.7. The estimates of the coefficients in Γ, similar to those presented in Table 2.6, are left untabulated to ease the presentation.

I find strong evidence about portfolio trading being correlated with asymmetric pricing. When customers buy portfolios from dealers, they pay 13.34 bps less for each bond compared to what they would pay when buying them sequentially. In turn, when customers sell portfolios to dealers, they pay 3.09 bps more than when doing it sequentially. These numbers represent a 42.6% discount when buying and a 9.9% penalty when selling portfolios, respectively. The asymmetric coefficients hold robustly when I estimate all the alternative model specifications described when presenting Table 2.6.

Table 2.7: Transaction costs regression on trade characteristics and trade side.

Dependent Variable:	Transaction Cost					
	Baseline	Dealer FE	Bond FE	No DST	No Offset	No Mixed
Portfolio	-13.34***	-10.78***	-10.92***	-13.45***	-14.32***	-13.16***
	(0.92)	(0.79)	(0.85)	(1.01)	(0.94)	(1.02)
Customer Sell	-9.54***	-6.94***	-9.13***	-9.55***	-9.00***	-9.55***
	(0.47)	(0.43)	(0.43)	(0.47)	(0.48)	(0.47)
Port. $\times$ Cust. Sell	16.43***	13.34***	15.96***	17.99***	16.92***	18.69***
	(1.53)	(1.42)	(1.46)	(1.72)	(1.56)	(1.87)
$\beta_1 + \beta_3$	3.09**	2.56**	5.04***	4.54***	2.6**	5.53***
	(1.22)	(1.11)	(1.12)	(1.38)	(1.24)	(1.56)
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Issuer Industry FE	Yes	Yes	No	Yes	Yes	Yes
Dealer FE	No	Yes	No	No	No	No
Bond FE	No	No	Yes	No	No	No
Observations	6,300,985	6,300,985	6,307,999	6,279,622	6,021,275	6,276,809
Adjusted R^2	0.098	0.204	0.144	0.098	0.113	0.098
Within R^2	0.096	0.032	0.034	0.096	0.110	0.096

Note: This table provides OLS estimates of the trade-level Equation (2.3). The baseline specification regresses transaction cost on a portfolio trade dummy, a customer sell dummy, the interaction of the portfolio trade and customer sell dummies, a portfolio dealer dummy, age, amount outstanding, time to maturity, credit rating, trade size, day fixed effects and issuer industry fixed effects. Alternative specifications include dealer fixed effects (column 2), bond fixed effects (column 3), the exclusion of portfolios executed within a 5-minute window of delayed spot times (column 4), the exclusion of "Offset$\leq 15m$ - C" trades (column 5), and the exclusion of mixed portfolios (column 6). To ease the exposition, some estimates are left untabulated. Clustered day-bond standard errors are shown in parentheses. One, two, and three stars indicate statistical significance at the 0.1, 0.05, and 0.01, respectively.

The evidence in Table 2.7 suggests that large balance sheet expansions may be playing a role when dealers price portfolios, as incoming portfolios are penalized. These results are in sharp contrast with those found in previous studies (Meli and Todorova, 2022; Li, O'Hara, Rapp, and Zhou, 2023), where portfolio trading is consistently less expensive than sequential trading. In the next section, I formally study the alternative drivers behind the found discounts and penalties.

2.5 Transaction Costs Drivers

To study what drives the differences in transaction costs between portfolio and sequential trading, I proceed in two steps. Firstly, I address how individual bonds are priced within the portfolios. This analysis answers questions regarding whether some segments of the market, e.g. risky bonds or small issues, are driving the effects seen in subsection 2.4.2. I find a significant cross-subsidy within portfolios: characteristics that are priced in sequential trading are reversed when the bond is included in a portfolio. Secondly, I investigate what portfolio characteristics are priced by dealers and in which direction. I enhance the trade-level estimations using portfolio characteristics and find significant evidence of both balance sheet effects and portfolio diversification effects.

2.5.1 Bonds Transaction Cost Drivers and Portfolio Trading

I start by extending the baseline Equation (2.2) interacting all variables in the vector C with the portfolio trading dummy. Since trades within portfolios are priced differently according to their side, I estimate this equation for buy trades and sell trades separately.

$$TC_i = \alpha + \beta \mathbf{1}_{i=\text{Portfolio}} + \Gamma_1 C_i + \Gamma_2 C_i \mathbf{1}_{i=\text{Portfolio}} + \Lambda FE + \epsilon_i, \tag{2.4}$$

The estimation results are presented in Table 2.8. To simplify the exposition, I present in the second and fourth columns the estimated coefficients of the interacted variables. As can be seen, there is a clear pricing reversal within portfolios. For example, bonds with high credit risk (CCC-D) are costly to trade when doing so sequentially, paying 44.2 bps and 31.4 bps more than low-credit-risk bonds (A-BBB). However, when those low-rated bonds are included in a portfolio, their pricing improves and the risk effect is partially canceled out. A similar pattern happens with virtually all variables included in vector C.

The observed price reversal is not surprising, as portfolios allow to diversify the risk implied by holding a single security. To deepen into this idea, I follow the long-standing Capital Asset Pricing Model tradition and compute what fraction of a bond (excess) returns variance is explained by factors other than markets' fluctuations (see Appendix 2.A.5). The higher this fraction is, the larger the diversification gains a bond inherits when it is included in a portfolio, thus we should expect large price reversals. Table 2.8 supports this hypothesis, with a full reversal for customer buys and a partial reversal for customer sells.

On top of estimating all the different specifications described in Tables 2.6 and 2.7, under which the price reversal holds robustly (untabulated), I perform one additional check specific to this result. Although dealers should report the price of each specific bond traded to FINRA, portfolios are traded at a unique price. Since the portfolio price is the one with economic significance, it may be the case that the individual prices reported for portfolio trading bonds are non-informative. Taking the argument to the limit, any vector of prices for which its (volume-weighted) sum equals the portfolio price could be reported. This would give room for a mechanical price reversal, in which all bond prices within a portfolio are reported to be equal. I discard such a claim relying on two facts. First, TRACE provides incentives for dealers to upload prices according to market valuation, regardless of the trading protocol used. Particularly, "*TRACE will validate the price that the user has submitted by comparing it to other recent transactions in the same security. If the reported price is substantially different than the price determined by TRACE to be the "current market" for that security, an error message will be generated.*".[14] Second, in Appendix 2.A.7, I show that the pricing of bond characteristics within portfolios follows the same patterns as in sequential trading, rejecting thus the hypothesis of a non-informative reported price vector.

[14]See TRACE User Guide 2023, p31.

Table 2.8: Transaction costs regression on interacted trade characteristics.

Dependent Variable:	Transaction Cost			
	Customer buys		Customer sells	
		× Portfolio		× Portfolio
Portfolio	41.80***		22.04***	
	(2.87)		-2.56	
Portfolio Dealer	-30.94***		-14.93***	
	(0.57)		-0.4	
Age	-0.39**	0.59***	0.73***	-1.02***
	(0.17)	(0.21)	(0.1)	(0.21)
Amount Outstanding	-0.82**	1.10**	-1.76***	-0.75*
	(0.37)	(0.48)	(0.3)	(0.40)
Time-to-maturity 1-3	-13.56***	12.80***	-4.56***	3.74***
	(0.72)	(1.12)	(0.88)	(1.26)
Time-to-maturity 5-10	16.75***	-16.92***	6.08***	-2.28**
	(0.88)	(1.12)	(0.6)	(0.96)
Time-to-maturity >10	60.05***	-49.47***	19.93***	-14.86***
	(2.08)	(3)	(0.97)	(3.07)
Micro (<100K)	24.49***	-25.34***	11.54***	-8.07***
	(0.54)	(1.16)	(0.38)	(1.48)
Round (1M-5M)	-12.83***	19.93***	-4.72***	8.84***
	(0.58)	(1.48)	(0.47)	(1.93)
5M and above	-10.12***	34.24***	0.85	11.45***
	(0.76)	(9.7)	(0.56)	(2.46)
IG (AAA-AA)	-6.50***	5.89***	-1.39***	2.60**
	(0.76)	(1.29)	(0.45)	(1.30)
HY (BB-B)	20.85***	-19.21***	9.50***	-8.61***
	(1.11)	(1.6)	(0.65)	(1.64)
HY (CCC-D)	44.20***	-36.75***	31.38***	-26.73***
	(4.21)	(5.04)	(3.69)	(4.67)
Idiosync. var. share	35.89***	-37.30***	27.42***	-13.48***
	(2.81)	(3.46)	(2.07)	(3.09)
Day FE	Yes		Yes	
Issuer Industry FE	Yes		Yes	
Observations	3,814,350		2,349,074	
Adjusted R^2	0.145		0.051	
Within R^2	0.142		0.046	

Note: This table provides OLS estimates of the trade-level Equation (2.4). Transaction cost is regressed on a portfolio trade dummy, a portfolio dealer dummy, trade characteristics –age, amount outstanding, time to maturity, credit rating, trade size, and idiosyncratic variance share–, the interaction of trade characteristics and the portfolio trade dummy, day fixed effects and issuer industry fixed effects. Equation (2.4) is estimated for customer buy trades and customer sell trades separately. Columns 2 and 4 show the estimates for the interacted trade characteristics. Clustered day-bond standard errors are shown in parentheses. One, two, and three stars indicate statistical significance at the 0.1, 0.05, and 0.01, respectively.

2.5.2 Portfolios Transaction Cost Drivers

Once shown that the characteristics that drive transaction costs in sequential trading are partially reversed when bonds are traded through portfolios, I proceed to address what portfolio characteristics determine its transaction costs. I expand Equation (2.2) decomposing the portfolio dummy into a vector that locates portfolios into several categories:

$$TC_{i,p} = \alpha + \beta \mathbf{1}_{i=\text{Portfolio}} + \Gamma C_i + \Delta \mathbf{1}_{i=\text{Portfolio}} D_p + \Lambda FE + \epsilon_{i,p}, \tag{2.5}$$

where vector D includes portfolio characteristics: number of bonds, volume Herfindahl-Hirschman Index (HHI), credit rating average, standard deviation compared to its i.i.d. counterfactual, amount outstanding average, and aggregate volume (see Appendix 2.A.5). Except for the HHI, the remaining variables are incorporated as dummies that indicate if a portfolio belongs to a specific bin regarding quartile partitions.

The set of portfolio variables aims to cover alternative hypotheses that may drive dealers to charge customers different prices when trading portfolios than when trading those bonds sequentially. First, the aggregate volume of each portfolio tells us how much balance sheet space a dealer needs to incur, thus addressing directly the balance sheet channel. Second, I use several variables that indirectly measure the gains from risk diversification a portfolio can provide. The variance of portfolio returns mechanically decreases in the number of bonds and increases when portfolio weights are concentrated, the latter considering a scenario where all bonds have similar individual variances. Additionally, when the average credit rating is high, there is more room for portfolios to diversify away the default risk. Finally, I compute the ratio between the return volatility of the portfolio and the one it would have should all the bonds in it be independently distributed. The smaller this ratio the higher the gains from diversification. The last channel tested is the asymmetric information channel: dealers may penalize portfolios when they infer that customers have private information about one or many bonds in the portfolio. I use the average amount outstanding as a proxy of customers' (lack of) private information, as larger bonds tend to have a wider investor base Brugler, Comerton-Forde, and Martin (2022). To further investigate the asymmetric information channel, I also estimate alternative equations where I exploit time-series information and the ex-post performance of the bonds traded.

Equation (2.5) is estimated for full buy, full sells, and mixed portfolios separately. In each case, portfolio trading bonds are compared with sequential buys, sells, and buys and sells, respectively. Again, the coefficients associated with bond characteristics in C are not shown to ease the exposition. Table 2.9 presents the estimation results.

As can be seen in Table 2.9, the balance sheet and diversification channels are economically and statistically significant. Portfolios that involve larger volumes pay higher transaction costs. Compared to those bonds in portfolios below the 25th percentile of the aggregate volume distribution, bonds in portfolios above the 75th percentile pay 36.34 bps, 9.11 bps, and 25.29 bps more transaction costs, according to the trade side considered. Since almost all bonds traded through portfolios imply balance sheet costs, the larger those costs the larger the transaction costs dealers translate to customers. Table 2.9 also shows that, for portfolios where customers buy bonds, the transaction costs are reduced as we increase the number of bonds. In particular, bonds in full customer-buy portfolios in the 4th quartile pay 27.67 bps less transaction costs than those in the 1st quartile, with a similar pattern happening for mixed portfolios. The other variables considered to address portfolio diversification present no clear evidence in favor or against the hypothesis.

I do not find strong evidence about asymmetric information driving portfolio transaction costs. The estimated coefficients associated with the average amount outstanding of a portfolio go in opposite directions according to buy and sell trades and are typically not

Table 2.9: Transaction costs regression on portfolio characteristics.

		Cust. Buy	Cust. Sell	Mixed
	Port. × Volume 25-50 pctl	16.89***	5.66*	2.72
		(2.10)	(3.00)	(1.96)
Balance Sheet	Port. × Volume 50-75 pctl	24.78***	6.61**	9.59***
		(2.71)	(2.97)	(2.30)
	Port. × Volume 75-100 pctl	36.34***	9.11**	25.29***
		(3.31)	(4.30)	(4.22)
	Port. × # Bonds 25-50 pctl	-9.11***	-0.11	-8.30***
		(2.48)	(2.48)	(2.17)
	Port. × # Bonds 50-75 pctl	-17.45***	8.02**	-12.82***
		(3.67)	(3.56)	(3.45)
	Port. × # Bonds 75-100 pctl	-27.67***	-5.74	-21.60***
		(4.98)	(6.48)	(5.27)
	Port. × HHI	-104.45	-53.22	-35.16**
		(66.14)	(44.38)	(16.76)
Risk Diversification	Port. × Avg Rating 25-50 pctl	-1.06	4.28	-0.76
		(2.93)	(2.87)	(3.46)
	Port. × Avg Rating 50-75 pctl	-5.25	-6.69*	-10.53**
		(3.26)	(3.63)	(4.27)
	Port. × Avg Rating 75-100 pctl	-8.66***	-5.37	-9.52**
		(3.25)	(3.63)	(4.13)
	Port. × SD/SDiid 25-50 pctl	2.76	-1.44	4.50**
		(2.42)	(2.78)	(1.97)
	Port. × SD/SDiid 50-75 pctl	0.39	-2.27	1.77
		(3.84)	(4.47)	(2.29)
	Port. × SD/SDiid 75-100 pctl	-5.67	4.82	1.76
		(4.99)	(7.61)	(2.90)
	Port. × Amount Outs. 25-50 pctl	1.91	1.53	4.08*
		(2.29)	(3.23)	(2.35)
Asymmetric Information	Port. × Amount Outs. 50-75 pctl	4.44**	-4.85*	-1.19
		(2.14)	(2.88)	(3.18)
	Port. × Amount Outs. 75-100 pctl	4.27	-0.36	-2.89
		(3.26)	(4.41)	(3.90)
Day FE		Yes	Yes	Yes
Issuer Industry FE		Yes	Yes	Yes
Observations		3,890,730	2,384,982	6,230,081
Adjusted R^2		0.141	0.049	0.095
Within R^2		0.138	0.045	0.093

Note: This table provides OLS estimates of the trade-level Equation (2.5). Transaction cost is regressed on a portfolio trade dummy, a portfolio dealer dummy, age, amount outstanding, time to maturity, credit rating, trade size, and the interaction of the portfolio trade dummy with portfolio characteristics –number of bonds, volume HHI, credit rating average, standard deviation compared to its i.i.d. counterfactual, amount outstanding average, aggregate volume–, day fixed effects and issuer industry fixed effects. Clustered day-bond standard errors are shown in parentheses. One, two, and three stars indicate statistical significance at the 0.1, 0.05, and 0.01, respectively.

significant. As the lack of significance may be due to the variable considered not being an accurate proxy for asymmetric information, in Appendix 2.A.8 I estimate two alternative models that speak to this channel. First, I extend Equation (2.2) by replacing the day fixed effects for time-series variables. Among them, the Volatility Index produced by the Chicago Board Options Exchange (VIX) measures the uncertainty related to stock price movements. If the asymmetric information channel plays a role when pricing portfolios, it is expected that such a role gains importance in uncertain times. I find no evidence regarding this claim. Second, I compute the evolution of bond prices after these had been traded, at different horizons. If portfolios are traded on information, it is expected that the prices of those portfolio bonds sold (bought) would decrease (increase) after the trade more than what they do after sequential trades (Di Maggio, Franzoni, Kermani, and Sommavilla, 2019; Pinter, Wang, and Zou, 2024). Again, I find no evidence supporting this hypothesis.

Overall, the evidence presented suggests that dealers price portfolios differently according to the aggregated volume traded and the amount of risk they can diversify. Larger portfolios imply higher balance sheet costs and thus are traded with a penalty. In turn, conditional on the aggregated volume traded, portfolios with more bonds reduce their return volatility and thus are traded with a discount.

2.6 Conclusion

This paper empirically studies portfolio trading in the corporate bond market. This new protocol allows customers to trade a bundle of bonds simultaneously, reducing the time it would take to trade these bonds sequentially and the consequent execution uncertainty. In line with the novelty of the protocol, data sets do not explicitly account for it. To overcome this issue, I develop an algorithm to infer portfolios from specific characteristics of bundles of bonds. I find that portfolio trades represent a significant and growing fraction of the market and that its intermediation is concentrated among top dealers, who source bonds using their balance sheets. Finally, I turn to the liquidity implications of portfolio trading. I do so by comparing the transaction costs charged in this protocol and in the alternative one, i.e. traditional sequential trading. I present novel evidence of asymmetrical transaction costs: compared to sequential trading, portfolio trading is 42.6% less expensive when customers buy and 9.9% more expensive when they sell. To address which factors drive these results, I proceed in two steps. On the one hand, I show there is a significant cross-subsidy within portfolios: bond characteristics that are priced in sequential trading are reversed when the bond is included in a portfolio. On the other hand, I study several hypotheses of portfolio pricing. I find that dealers penalize portfolios that involve large balance sheet costs and offer discounted transaction costs to those portfolios that diversify risk. I find no evidence of asymmetric information driving portfolio pricing.

2.A Appendix

2.A.1 Customer-dealer trades frequency

Here I show that bundles of 30 or more bonds being traded by the same dealer at the same second are rare, a fact that supports my portfolio identification strategy. Table 2.A.1 presents statistics for the top ten dealers performing portfolio trades. Both taking into account the extensive margin, i.e. how often a customer-dealer trade is observed, and the intensive margin, i.e. how many customer-dealer trades happen in every trading second, it is observed that trading is rather infrequent, with bundles of 30 or more bonds only observed at the extreme tail of the distribution.

Table 2.A.1: Customer-dealer trades frequency, per portfolio dealer. Period 2018-2019.

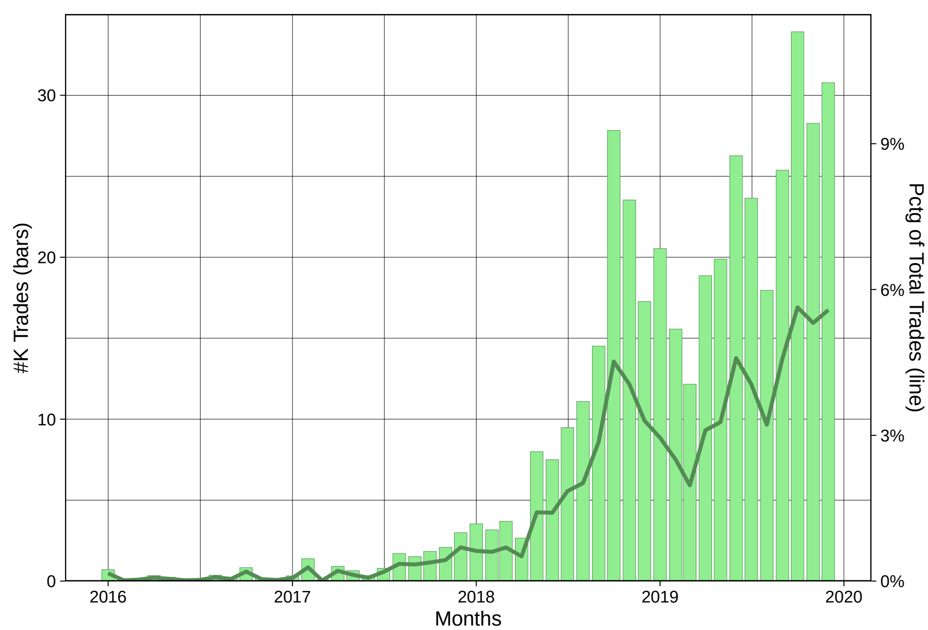
	Vol % share		Seconds between trades in an hour				Number of trades in a second			
Dealer	Port.	Seq.	p50	p90	p99	p99.9	p50	p90	p99	p99.9
1	49.9	10.2	51	26	21	16	1	3	6	57
2	18.6	8.9	65	34	26	19	1	2	4	38
3	17.4	0.7	157	45	33	20	1	3	10	147
4	6.5	8.1	86	45	36	24	1	2	4	15
5	3.3	8.5	97	51	38	21	1	2	4	10
6	2.5	7.5	90	40	31	20	1	3	6	10
7	0.7	8.2	103	53	40	19	1	2	4	10
8	0.2	5.3	138	73	56	43	1	1	2	5
9	0.2	0.0	1,800	240	93	67	1	9	11	69
10	0.2	0.3	720	95	59	48	1	6	14	28

Note: This table shows statistics for the top ten portfolio trading dealers. Columns 2 and 3 show the market share of each dealer, for portfolio and sequential trading, respectively. Columns 4-7 measure how often a customer-dealer trade is observed. To compute this variable, I initially calculate how many customer-dealer trades a dealer executes in every hour in which she executes a trade (avoiding thus the hours in which there is no market). Then I divide 3600 by such a figure to re-express the variable as the number of average seconds between trades in each hour. For example, if in an hour there are 10 trades, that means that a trade happens on average every 3600/10=360 seconds during that hour. Columns 8-11 measure how many customer-dealer trades happen in every trading second.

2.A.2 Portfolio Trades Market Share

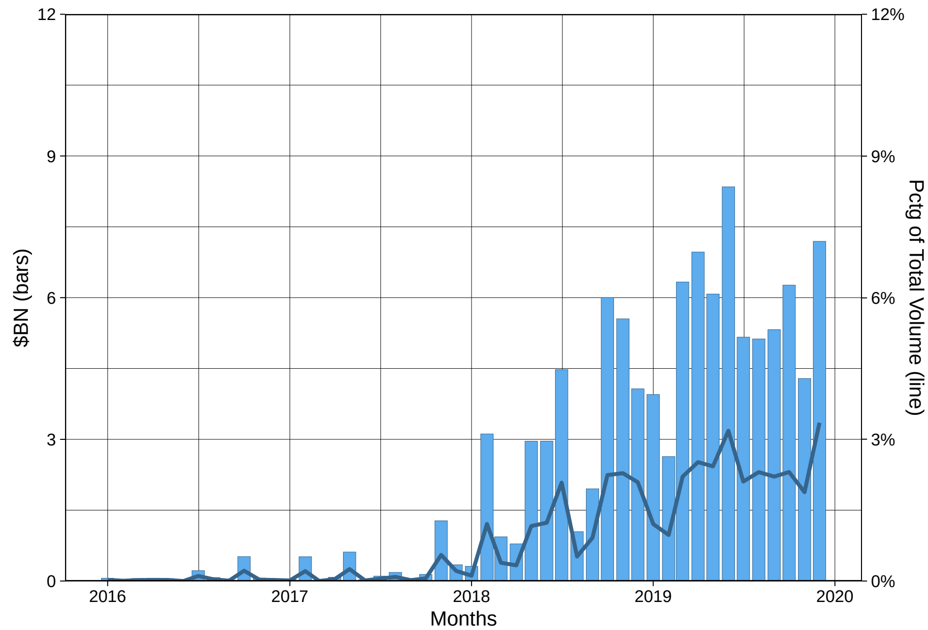
Here I present the monthly time series of portfolio trading shares considering both volume and the amount of trades, for alternative market segments.

Figure 2.A.1: Portfolio trading trades - All segments.



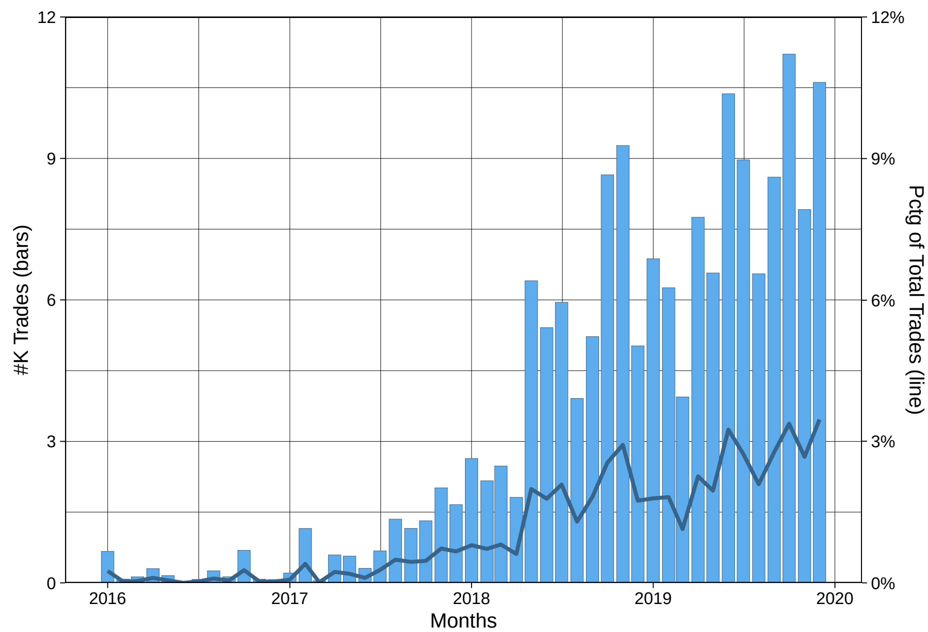
Note: This figure depicts the monthly time-series of trades performed through portfolio trading, including both customer-dealer and inter-dealer trades. The bars –left axis– indicate the number of trades, expressed in thousands. The line –right axis– indicates market share, expressed in percentage points.

Figure 2.A.2: Portfolio trading volume - Customer dealer segment.



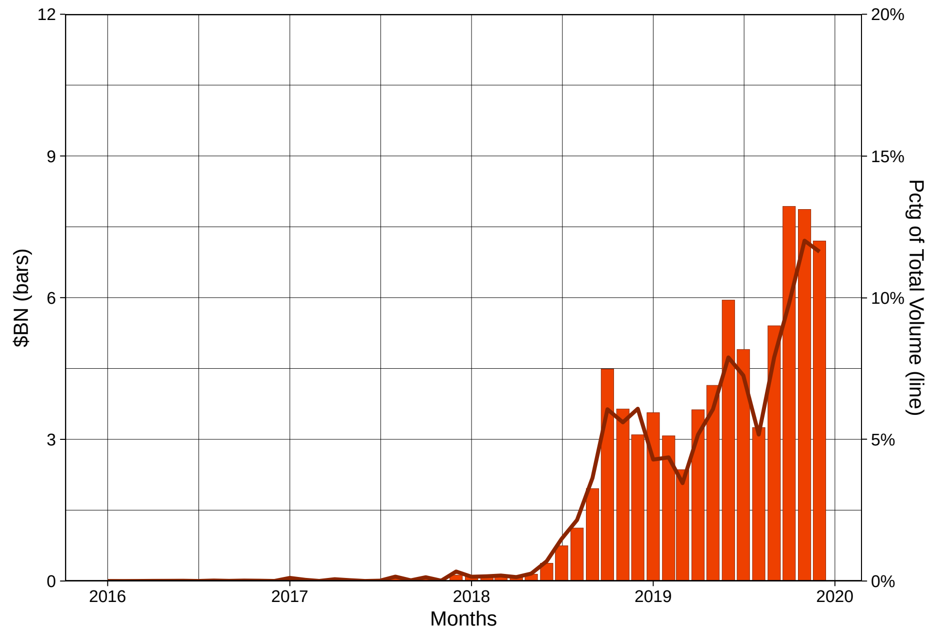
Note: This figure depicts the monthly time-series of portfolio trading volume, including only customer-dealer trades. The bars –left axis– indicate total face value, expressed in billion dollars. The line –right axis– indicates market share, expressed in percentage points.

Figure 2.A.3: Portfolio trading trades - Customer dealer segment.



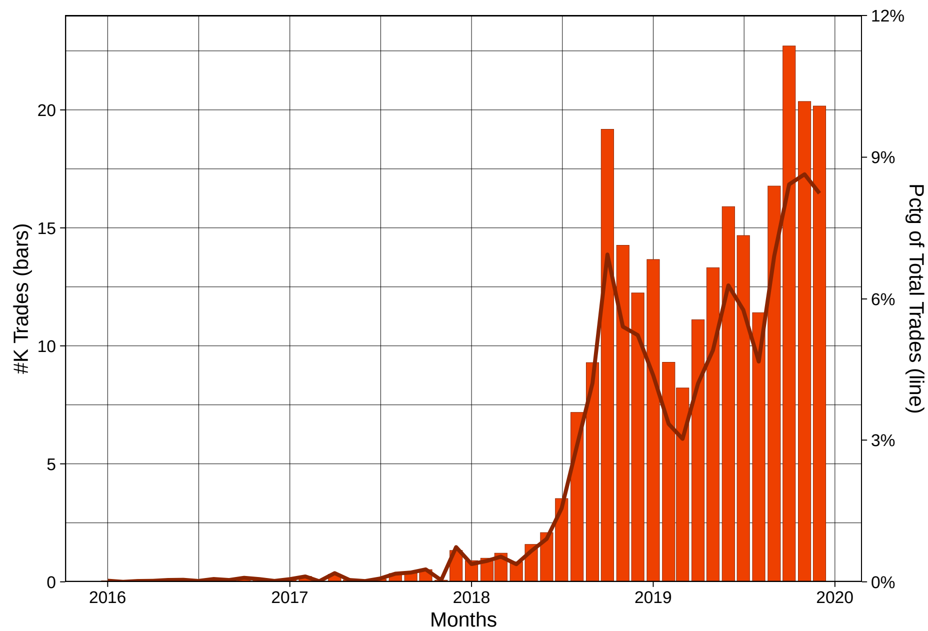
Note: This figure depicts the monthly time-series of trades performed through portfolio trading, including only customer-dealer trades. The bars –left axis– indicate the number of trades, expressed in thousands. The line –right axis– indicates market share, expressed in percentage points.

Figure 2.A.4: Portfolio trading volume - Inter-dealer segment.



Note: This figure depicts the monthly time-series of portfolio trading volume, including only inter-dealer trades. The bars –left axis– indicate total face value, expressed in billion dollars. The line –right axis– indicates market share, expressed in percentage points.

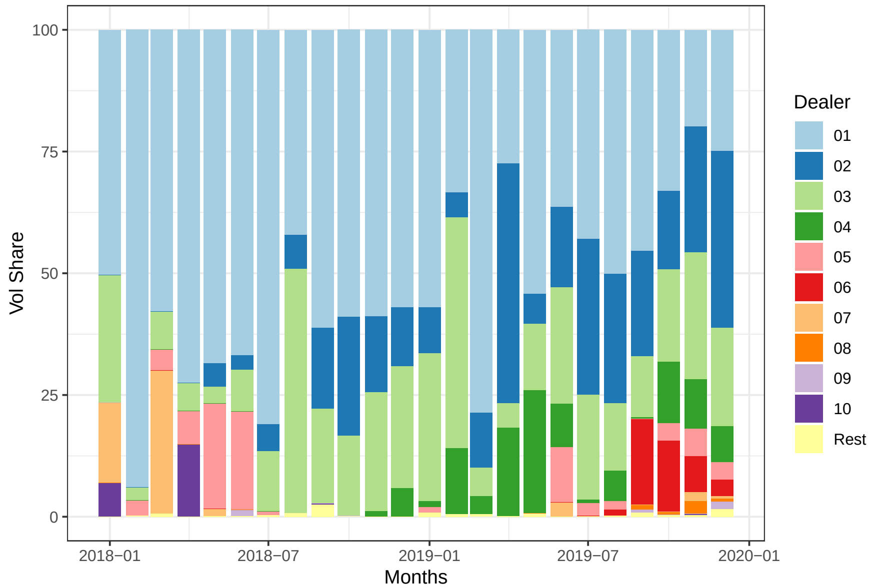
Figure 2.A.5: Portfolio trading trades - Inter-dealer segment.



Note: This figure depicts the monthly time-series of trades performed through portfolio trading, including only inter-dealer trades. The bars –left axis– indicate the number of trades, expressed in thousands. The line –right axis– indicates market share, expressed in percentage points.

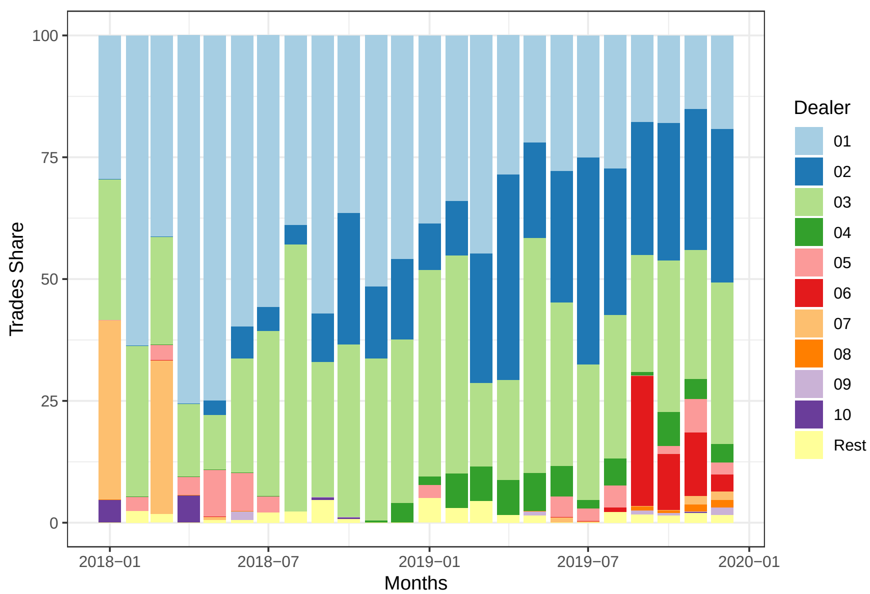
2.A.3 Dealers' Market Share Evolution of Portfolio Trading

Figure 2.A.6: Portfolio Trading Market Share Evolution - Volume.



Note: This figure depicts dealers' monthly share of the portfolio trading (face value) volume. Dealers are ordered according to their volume share in the entire period 2018-2019.

Figure 2.A.7: Portfolio Trading Market Share Evolution - Trades.



Note: This figure depicts dealers' monthly share of the portfolio trading trades. Dealers are ordered according to their volume share in the entire period 2018-2019.

2.A.4 Bonds Sourcing using Number of Trades

Table 2.A.2 shows how the top ten portfolio dealers source their portfolio and sequential trades: offsetting with other dealers or customers, or involving their own inventories. The figures express percentage points computed out of the number of trades.

Table 2.A.2: Sourcing of Portfolio - Number of Trades.

	Market Share		Portfolio Sourcing			Sequential Sourcing		
			Offset ≤15m		Non-Offset	Offset ≤15m		Non-Offset
Dealer	Portfolio	Sequential	C	D		C	D	
1	35.3	6.7	2.6	0.4	97.0	4.0	7.6	88.3
2	21.2	4.3	2.0	1.0	97.0	5.5	2.2	92.3
3	31.0	1.7	0.0	1.1	98.9	0.0	1.1	98.9
4	3.2	3.1	2.4	1.2	96.4	6.0	17.2	76.8
5	2.3	2.9	14.1	0.1	85.8	9.4	1.2	89.5
6	3.3	4.0	0.6	0.4	99.1	4.5	1.3	94.2
7	1.5	2.7	0.2	0.2	99.6	8.7	1.6	89.7
8	0.3	1.8	0.2	0.5	99.3	10.0	5.4	84.6
9	0.3	0.1	0.0	98.5	1.5	0.2	72.5	27.3
10	0.2	0.3	0.0	100.0	0.0	0.0	100.0	0.0

Note: This table shows, for each of the top ten portfolio trading dealers, its portfolio trading market share (column 2), its sequential trading market share (column 3), the distribution in the three categories – Offset ≤ 15 - C, Offset ≤ 15 - D, Non-Offset – of its portfolio trading activity (columns 4-6) and sequential trading activity (columns 7-9). All statistics are computed using the non-weighted number of trades.

2.A.5 Variables Computation

Trade-level variables:

- Portfolio dealer: Dummy variable that equals 1 if the trade was performed by a dealer that accumulates more than 0.01% of the total portfolio trading volume.
- Age: Number of years between the day of offering and the trading day.
- Amount Outstanding: Total amount outstanding of the bond being traded, measured in face value and expressed in billions of dollars.

- Time to Maturity: Number of years between the day of maturity and the trading day.

- Trade Size: Par-value of the transaction, expressed in millions of dollars.

- Credit Rating: I initially compute the average letter ratings of the three agencies present in FISD (S&P, Moodie's, and Fitch) by using standard letter-number equivalences (e.g., AAA=1, D=25). I then go back to letter ratings using the same equivalence and classify bonds as Investment Grade Superior, Investment Grade Inferior, High Yield Superior, or High Yield Inferior if they belong to credit rating brackets AAA-AA, A-BBB, BB-B, or CCC-D, respectively.

- Turnover: I compute the turnover of a bond over the last 3 months previous to the month in which it is traded. For each bond, past turnover equals $\sum_{s=1}^{s=3} vol_{t-s}/(\sum_{s=1}^{s=3} iao_{t-s})/3)$, where t is the month in which the trade happens, vol_{t-s} is the total face value traded in month $t-s$, and iao_{t-s} is the mean amount outstanding during month $t-s$.

- Idiosyncratic variance share: I firstly compute bond i weekly returns $R_{i,w}$ using volume-weighted average prices, including accrued interest rates and coupon payments. Second, I compute the OLS residuals of the regression $R_{i,w} - R_w^f = \alpha + \beta(R_w^m - R_w^f) + \epsilon_{i,w}$, where R_w^f is the weekly interpolated 1M Treasury rate and R^m is the weekly return of the Bank of America Merrill Lynch US Corporate Index (IG or HY according to the bond considered). Finally, I compute the idiosyncratic variance share as the ratio $Var(\hat{\epsilon}_{i,w})/Var(R_{i,w} - R_w^f)$. This variable is only computed for those bonds with at least 30 weekly returns.

Portfolio-level variables:

- Number of bonds: Sum of bonds in a portfolio

- Herfindahl-Hirschman Index (HHI): $\sum_{i\in p}(\text{vol}_i/\sum_{i\in p}\text{vol}_i)^2$, where vol_i denotes the Trade Size of trade i in portfolio p.

- Average Rating: Simple average of the Credit Rating of the bonds in a portfolio, where the Credit Rating character variable is turned to numeric by using standard letter-number equivalences (e.g., AAA=1, D=25).

- Portfolio Standard Deviation compared to its iid counterfactual (SD/SDiid): I initially compute the portfolio return standard deviation SD. For this, I take bond returns $R_{i,w}$ as previously described and impute weights W_i using the net volume (face value) of bonds in the portfolio. Secondly, I compute the counterfactual iid portfolio return standard deviation $SD_{iid} = [\sum_i w_i^2 Var(R_i)]^{1/2}$. Finally, I compute the percentage deviation and express it in percentage points $100(SD/SD_{iid} - 1)$. This variable is only computed using those bonds that, in the previous 30 weeks before the portfolio was traded, have at least 15 weekly returns computed.

- Amount Outstanding: Simple average of the Amount Outstanding of the bonds in a portfolio.

- Volume: Sum of Trade Size of the bonds in a portfolio.

2.A.6 Subsample of Customer-Dealer Trades with Reference Price Available

To construct the transaction cost measure for customer-dealer trades, there should exist at least one same bond-day inter-dealer trade from which to take the reference price. In this Appendix, I present how this requirement reduces the portfolio and sequential trading subsamples.

Table 2.A.3 shows how the overall number of observations and volume implied is reduced when we only consider those customer-dealer trades with an associated reference price. The reduction is higher in the portfolio trade subsample.

Table 2.A.3: Trades with associated reference price.

Sample	Observations (%)	Volume (%)
Portfolio	60.03	62.86
Sequential	83.10	69.88

The reduction in the samples that are used for the transaction costs analysis can represent a concern if the lack of reference price correlates with trade characteristics. In such a case, our estimations may suffer from a selection bias. Tables 2.A.4 and 2.A.5 decompose portfolio and sequential samples into those trades with and without an associated reference price, and present the distribution of relevant characteristics in the two partitions. Although there are clear differences between the partitions with and without a reference price, we still have enough variation in each characteristic so that we can control for them in the estimations, thus lessening the selection bias concern.

Table 2.A.4: Variables distribution differences within portfolio trades.

Variables	Ref Price	Mean	Std. dev.	.05	.25	.50	.75	.95
Age (years)	No	3.27	3.11	0.35	1.15	2.38	4.41	9.09
	Yes	3.23	2.75	0.36	1.32	2.63	4.41	7.63
Amount Outs. $B	No	0.80	0.57	0.30	0.50	0.64	1.00	1.80
	Yes	1.25	1.00	0.40	0.62	1.00	1.50	3.00
Customer Sell	No	0.43	0.50	0.00	0.00	0.00	1.00	1.00
	Yes	0.41	0.49	0.00	0.00	0.00	1.00	1.00
Maturity (years)	No	11.18	9.53	2.48	4.76	6.99	17.95	28.93
	Yes	8.81	7.77	2.18	4.30	6.33	8.63	27.85
Rating 1-25	No	10.81	3.78	5.00	8.00	11.00	14.00	17.00
	Yes	10.79	3.73	5.00	8.00	11.00	13.00	16.00
Trade Size $M	No	0.61	1.77	0.02	0.10	0.20	0.50	2.19
	Yes	0.69	2.00	0.02	0.10	0.25	0.50	2.50
Turnover 3m	No	21.42	42.02	2.85	9.47	17.27	27.51	52.91
	Yes	27.58	38.40	6.32	13.83	22.16	34.23	66.82

Table 2.A.5: Variables distribution differences within sequential trades.

Variables	Ref Price	Mean	Std. dev.	.05	.25	.50	.75	.95
Age (years)	No	3.40	3.18	0.33	1.13	2.51	4.70	9.00
	Yes	4.01	3.34	0.47	1.79	3.31	5.38	8.82
Amount Outs. \$B	No	0.81	0.64	0.28	0.45	0.60	1.00	2.00
	Yes	1.27	1.21	0.28	0.50	1.00	1.50	3.10
Customer Sell	No	0.50	0.50	0.00	0.00	1.00	1.00	1.00
	Yes	0.38	0.49	0.00	0.00	0.00	1.00	1.00
Maturity (years)	No	11.29	10.04	1.85	4.30	6.93	18.98	29.24
	Yes	7.39	7.40	1.40	3.05	5.15	7.86	26.30
Rating 1-25	No	9.69	3.77	5.00	7.00	9.00	12.00	17.00
	Yes	9.09	3.64	4.00	7.00	9.00	11.00	16.00
Trade Size \$M	No	1.39	3.08	0.01	0.07	0.30	1.42	5.91
	Yes	0.66	2.32	0.01	0.02	0.05	0.25	3.50
Turnover 3m	No	19.70	24.07	2.37	7.78	14.57	25.11	53.79
	Yes	23.15	23.84	4.06	10.19	16.60	27.93	64.97

2.A.7 Transaction Costs Drivers Within Portfolios

In this Appendix I provide evidence on individual reported prices of portfolio trading bonds being economically significant. I do so by showing that the pricing of bond characteristics within portfolios follows the same patterns as in sequential trading. Using only the portfolio trading observations, I estimate the following equation:

$$TC_{i,p} = \alpha + \Gamma C_i + \delta FE_p + \Lambda FE + \epsilon_{i,p},$$

where I include portfolio fixed effects to capture how characteristics included in vector C are priced within each portfolio. Table 2.A.6 shows the same pricing pattern as in sequential trading: smaller issues, with higher time to maturity and worse credit risk are more expensive to trade. These results hold under an alternative specification in which, instead of using portfolio fixed effects, I re-compute variables as quartile bins for each portfolio.

Table 2.A.6: Transaction costs regression on trade characteristics within portfolios.

Dependent Variable:	Transaction Cost
	(1)
Age	0.15*
	(0.08)
Amount Outstanding	-0.65***
	(0.17)
Time-to-maturity 3-5	0.54
	(0.54)
Time-to-maturity 5-10	1.18*
	(0.63)
Time-to-maturity >10	5.77***
	(1.48)
Odd (100K-1M)	0.23
	(0.49)
Round (1M-5M)	0.63
	(0.81)
5M and above	3.87**
	(1.71)
IG (A-BBB)	-0.16
	(0.56)
HY (BB-B)	2.21**
	(1.12)
HY (CCC-D)	7.01***
	(1.83)
Idiosync. var. share	1.41
	(1.17)
Customer Sell	3.79**
	(1.90)
Day FE	Yes
Portfolio FE	Yes
Issuer Industry FE	Yes
Observations	89,104
Adjusted R^2	0.134
Within R^2	0.003

Note: This table provides OLS estimates of the trade-level regression of transaction cost on age, amount outstanding, time to maturity, credit rating, trade size, idiosyncratic variance share, day fixed effects, portfolio fixed effects and issuer industry fixed effects. The sample consists of portfolio trades. Clustered day-bond standard errors are shown in parentheses. One, two, and three stars indicate statistical significance at the 0.1, 0.05, and 0.01, respectively.

2.A.8 Asymmetric Information Channel Robustness Checks

In this Appendix, I provide two alternative model specifications searching for evidence of an asymmetric information channel in portfolio transaction costs. The first model provided is an extension of Equation (2.2) in which I replace day fixed effects for time-series variables. This model allows me to include the Volatility Index (VIX), which is a time-series measure of market uncertainty. If the asymmetric information channel plays a role when pricing portfolios, it is expected that such a role gains importance in uncertain times.

$$\begin{aligned} TC_{i,t} =& \alpha + \beta_1 \mathbf{1}_{i=\text{Portfolio}} + \gamma C_i + \beta_2 \text{VIX}_t + \beta_3 \mathbf{1}_{i=\text{Portfolio}} \text{VIX}_t \\ &+ \beta_4 \text{T2Y-T1M} + \beta_5 \text{TED Spread} + \Lambda FE + \epsilon_{i,t} \end{aligned}$$

In Table 2.A.7, I estimate the model for customer buy and customer sell bonds separately. To control for the time-varying financial costs of dealers, I include the difference between the 2-year and 1-month Treasury rates (T2Y-T1M) and the difference between the 3-month LIBOR rate and 3-month Treasury rate (TED Spread). In columns 1 and 2 I estimate the model using bond fixed effects, while in columns 3 and 4 I use the vector C of bond characteristics plus dealer and industry fixed effects. The portfolio transaction costs differential with sequential trading does not change significantly in times of high expected volatility. This non-significance result holds if I control for VIX non-linearities by using quartile dummies.

For the second model, I compute the (ex-post) performance of bonds, at different horizons h (Di Maggio, Franzoni, Kermani, and Sommavilla, 2019; Pinter, Wang, and Zou, 2024):

$$\text{Performance}_{b,t,h} = [ln(P_{b,t+h}) - ln(P_{b,t})] * Q,$$

Table 2.A.7: Transaction Costs regression on time series macro variables.

Dependent Variable:	Transaction Cost			
	Customer Buy	Customer Sell	Customer Buy	Customer Sell
T2Y-T1M	5.82***	-2.91***	-0.51	-2.02***
	(0.61)	(0.60)	(0.53)	(0.54)
TED Spread	2.47	7.74***	3.25**	8.90***
	(1.79)	(1.74)	(1.47)	(2.24)
VIX	0.51***	0.14*	0.32***	0.15
	(0.05)	(0.08)	(0.03)	(0.11)
Portfolio × VIX	-0.32	0.38	-0.12	0.30
	(0.21)	(0.30)	(0.21)	(0.30)
Bond FE	Yes	Yes	No	No
Dealer FE	No	No	Yes	Yes
Issuer Industry FE	No	No	Yes	Yes
Observations	3,851,178	2,370,309	3,847,660	2,366,885
Adjusted R^2	0.197	0.082	0.283	0.109
Within R^2	0.036	0.006	0.057	0.012

Note: This table provides OLS estimates of the trade-level regression of transaction cost on a portfolio trade dummy, trade size, VIX, 2-year Treasury rate minus 1-month Treasury rate, TED Spread, and bonds fixed effects, for customer buy trades (column 1) and customer sell trades (column 2) separately. Alternatively, columns 3 and 4 replace bond fixed effects for age, amount outstanding, time to maturity, credit rating, dealer fixed effects and issuer industry fixed effects, for customer buy and customer sell trades, respectively. To ease the exposition, some estimates are left untabulated. Clustered day-bond standard errors are shown in parentheses. One, two, and three stars indicate statistical significance at the 0.1, 0.05, and 0.01, respectively.

where Q is a trade side indicator that equals 1 (-1) if the customer buys (sells) and $P_{b,t}$ is the simple average price of bond b at day t. In this way, each trade i will have attached a performance measure. Then I estimate a model where the performance attached to each trade is a function of its inclusion in portfolio trading, trade side, and relevant fixed effects.

$$\text{Performance}_{i,h} = \alpha + \beta_1 \mathbf{1}_{i=\text{Portfolio}} + \beta_2 \mathbf{1}_{i=\text{Cust. sells}} + \beta_3 \mathbf{1}_{i=\text{Portfolio}} \mathbf{1}_{i=\text{Cust. sells}} + \Lambda FE + \epsilon_{i,h}$$

If portfolios are traded on information, it is expected that the prices of those portfolio bonds sold (bought) would decrease (increase) after the trade more than what they do after sequential trades. Table 2.A.8 shows no evidence supporting this story. On the contrary, bonds sold through portfolios show a significant worse performance (price increase) than

Table 2.A.8: Return Performance regression on portfolio trading.

Dependent Variables:	Performance		
	h = 1 day	h = 10 days	h = 20 days
Portfolio	-0.89	-1.16	5.23
	(1.45)	(5.64)	(6.30)
Portfolio Dealer	6.23***	6.42***	5.93***
	(0.21)	(0.65)	(0.77)
Customer Sell	-6.41***	-10.43	-13.18
	(1.76)	(6.33)	(8.27)
Portfolio × Customer Sell	-3.22	-12.76	-32.87**
	(2.49)	(9.00)	(13.34)
Day FE	Yes	Yes	Yes
Bond FE	Yes	Yes	Yes
Observations	5,381,271	3,473,835	4,975,786
Adjusted R^2	0.020	0.019	0.020
Within R^2	0.002	0.001	0.001

Note: This table provides OLS estimates of the trade-level regression of performance on a portfolio trade dummy, customer sell dummy, the interaction between portfolio trade and customer sell dummies, trade size, day fixed effects, and bonds fixed effects. Estimates for the measured of performance at 1 day, 10 days, and 20 days horizons are presented in columns 1, 2, and 3, respectively. To ease the exposition, some estimates are left untabulated. Clustered day-bond standard errors are shown in parentheses. One, two, and three stars indicate statistical significance at the 0.1, 0.05, and 0.01, respectively.

those sold through sequential trading after 20 days of the trade.

Chapter 3

An International Macro Model with Convenience Yields

This paper presents a two-country model where the government bonds issued by one country can be used to ease financial transactions globally, resulting in endogenous convenience yields for these assets. I find that the new issuance of convenience assets spills over to foreign households, as their equilibrium transaction costs are reduced. Moreover, a global liquidity shock affects both countries differently, as the pricing of convenience assets increases in this shock and allows the issuing country to reduce taxes. Finally, I study the asset pricing implications of convenience yields in light of existing puzzles.

3.1 Introduction

In its pioneer work, Krishnamurthy and Vissing-Jorgensen (2012) defines a convenience asset as a security that provides relatively high liquidity and safety. Because of these "money like" features, investors are willing to accept a lower return than that of assets that do not to the same extent share these attributes. Particularly, the difference in interest rates that investors are willing to forego when holding a convenience asset is called the convenience yield. [1] [2]

The fact that convenience yields are particularly significant for dollar-denominated assets suggests that the U.S. occupies a central and distinctive role in international finance (Du, Im, and Schreger, 2018a; Liu, Schmid, and Yaron, 2020; Jiang, Krishnamurthy, and Lustig, 2020). Despite this role, the tradition in international economics is to model the U.S. and its trading counterparties as symmetric economies (Backus, Kehoe, and Kydland, 1992; Heathcote and Perri, 2014). What are the financial and real implications of the concentration of convenience assets in one country? Does the issuance of convenience assets spill over to foreign economies? Do countries' responses to global shocks differ when their debt is differently priced? What are the asset pricing consequences?

In this paper, I extend a canonical international macroeconomic model to allow for endogenous convenience yields. In particular, I consider the extreme case where only the home country issues convenience assets. I find that an increase in the issuance of convenience assets spills over to foreign households, as it relaxes their budget constraint by reducing transaction costs. Moreover, a global liquidity shock affects both countries differently, since the debt rollover cost for the convenience asset issuing country decreases and allows them to reduce taxes. Finally, I find that the inclusion of convenience yields helps match the theoretical predictions of the model with the empirical observations. In particular, it helps solve the risk-free rate puzzle, the risk premium puzzle, and the uncovered interest rate parity (UIP) puzzle.

The model developed is a two-country general equilibrium model *à la* Backus, Kehoe,

[1] The term convenience yield was originally used to refer to the interest rate that a hedged agent is willing to pay to borrow a storable commodity (Working, 1949).

[2] Convenience yields include but are not exhausted by liquidity premiums (Longstaff, 2004; Nagel, 2016; Lagos, Rocheteau, and Wright, 2017). Convenience assets are also extremely safe and thus are priced at a premium as they are used as collateral in financial transactions (Gorton, 2010), to back checkable deposits by commercial banks and money market funds (Bansal and Coleman, 1996), or to back long-term obligations (Greenwood and Vayanos, 2014).

and Kydland (1992) with endogenous convenience yields. In each country, denoted by home and foreign, consumption goods are produced using domestic and imported inputs. These goods are domestically absorbed by households, capital managers, and governments. The novel feature of the model is a consumption transaction cost function, which is decreasing in the amount of home bonds held by each household. This non-pecuniary value of home bonds generates endogenous convenience yields. Finally, the fiscal sector links the bond yields with taxes producing assymetric effects over households' budget constraints.

When the home country issues debt the equilibrium amount of convenience assets held by households increases and transaction costs are reduced. Therefore, an increase in the issuance of convenience debt has a spillover effect over the foreign country. Compared to the scenario without convenience yields, an increase in the home country's outstanding debt increases consumption in both countries. It also exacerbates the cross-country differences in output change: the increase in debt reduces the tax burden for the issuing country and, given local bias in the production of final goods, increases its output.

I also analyze the effect of a global liquidity shock. I do so by increasing the consumption transaction costs weight. In response to such a shock, households increase their demand for home bonds, as increasing the holdings of these assets reduces transaction costs and helps mitigate the negative effect of the liquidity shock. This *flight-to-liquidity* movement revaluates the home government's debt, allowing this government to reduce taxes. As can be seen, a liquidity shock in this environment implies a redistribution from the rest of the world to the convenience asset issuing country.

Finally, the inclusion of convenience yields helps to reconcile empirical patterns of rates of returns across assets and across countries with the theoretical predictions of the model. In compensation for reducing transaction costs, households are willing to forgo a fraction of the home bond's risk-free return, thus the equilibrium home risk-free rate is reduced. This wedge in the risk-free rate provides plausible equilibrium risk premiums without the need to impose an implausible risk-aversion parameter (risk premium puzzle; Weil, 1989) which would, in the absence of convenience yields, result in an implausible risk-free rate (risk-free rate puzzle; Mehra and Prescott, 1985). Since only the home bond returns include convenience yields, the model can account for the deviations observed between (same currency) risk-free rates of return across countries (uncovered interest rate parity puzzle; Fama, 1984). Consistent with the empirical literature, I show that these differences are stronger for short-term assets than for longer-term assets (Chinn, 2006; Du, Im, and Schreger, 2018a; Van Binsbergen, Diamond, and Grotteria, 2022).

3.1.1 Related Literature

This work is related to three strands of the literature. First, it is related to the literature on convenience yields (Krishnamurthy and Vissing-Jorgensen, 2012; Greenwood and Vayanos, 2014; Liu, Schmid, and Yaron, 2020). Particularly, it contributes to this line of work by introducing endogenous convenience yields into a two-country model in an asymmetric way. I extend the literature by studying the financial and real effects of issuing convenience assets, both domestically and abroad.

Secondly, I contribute to the international macroeconomics literature that highlights the central role of the U.S. in the "global financial cycle" (Rey, 2015). Broadly speaking, there are two main drivers of such a role. The first explanation states that foreign firm's currency mismatch – inflows in local currency and debt in U.S. dollars – expose foreign countries to large spillover effects from the U.S. monetary policy shocks (Ranciere, Tornell, and Vamvakidis, 2010; Rancière and Tornell, 2016; Jiang, Krishnamurthy, and Lustig, 2020). The second explanation depicts an international system where the U.S. can better manage risk and so it makes transfers to the rest of the world in turmoil times ("exorbitant duty") and collects an insurance premium in normal times ("exorbitant privilege") (Gourinchas and Rey, 2022; Maggiori, 2017). I propose a third channel, where U.S. government bonds are used to ease financial transactions internationally. When the U.S. expands its outstanding debt, foreign investors see their transaction costs reduced and so we observe real spillovers over the rest of the world.

Third, I contribute to the asset pricing literature, by revisiting some longstanding puzzles. I particularly revisit the risk premium puzzle (Weil, 1989), the risk-free rate puzzle (Mehra and Prescott, 1985), and the forward premium / UIP puzzle (Fama, 1984). I show that including convenience yields on government bonds introduces a wedge in the no-arbitrage asset pricing conditions, and thus we can obtain plausible values for risk premium and UIP deviations without imposing an unrealistic equilibrium risk-free rate.

Finally, a closely related work is Valchev (2020), which includes convenience yields in an international macro model. The author considers the symmetric case where both countries issue equally convenient assets. By restricting the issuance of convenience assets to only one country and linking government debt financing with fiscal policy, my model allows to study the cross-country real economy differences in the steady state and the asymmetric responses when countries suffer similar shocks.

The rest of the paper is organized as follows. Section 2 develops an international two-country model with convenience yields. Section 3 provides the parametrization used and presents impulse-response functions to understand how the inclusion of convenience assets affects the baseline model. Section 4 presents the results regarding asset pricing puzzles. Finally, Section 5 concludes.

3.2 The Model

In this section, I develop a two-country general equilibrium model with endogenous convenience yields. The model belongs to the Backus, Kehoe, and Kydland (1992) tradition (BKK from now on), particularly extending on Heathcote and Perri (2002). In each country, denoted by home and foreign, there is a representative firm that produces intermediate goods using local capital and local labor, and a representative firm that aggregates both local and imported intermediate goods to produce final goods. These final goods are locally absorbed by households, capital managers, and governments. The model features consumption transaction costs, which are decreasing in the amount of bonds held by each household. This non-pecuniary value of bonds generates endogenous convenience yields. Finally, a rich fiscal sector links convenience yields with real activity. In the following subsections, I describe the problem and the optimality conditions of each agent.

3.2.1 Intermediate Good Firms

Output Y_t is produced in each country by a representative firm through a Cobb-Douglas technology, using local labor, l_t, and local capital, K_t, as inputs. These firms are impacted by a country-specific stationary productivity process, denoted by ν_t, and a global non-stationary labor-augmenting process, denoted by Z_t:

$$Y_t = \exp(\nu_t) K_t^{\alpha} (Z_t l_t)_t^{1-\alpha}. \tag{3.1}$$

To make the model stationary, I de-trend all real variables except for labor, and denote them with lower-case letters. Taking capital return r_t and wages w_t as given, these firms minimize costs subject to the production technology described before. Here I state the problem faced by home firms, the foreign firms' problem is analogous.

$$\min_{\{k_t,l_t\}} \quad r_t k_t + w_t l_t$$
$$\text{s.t.} \quad y_t = \exp(\nu_t) k_t^{\alpha} l_t^{1-\alpha}.$$

The optimal levels of capital and labor are given by the first-order conditions (FOC):

$$r_t k_t = \alpha y_t,$$
$$w_t l_t = (1-\alpha) y_t.$$

I assume that the price $p_{a,t}$ at which home firms sell their output includes an exogenous markup of $\exp(\mu_t)$ over the marginal costs. Markups are modeled as country-specific stationary processes. Since the production technology faces constant return to scale, I can equate marginal costs with average costs and so $p_{a,t} y_t = \exp(\mu_t)[r_t k_t + w_t l_t]$. Hence, I re-write the FOC as:

$$r_t k_t = \alpha[r_t k_t + w_t l_t] = \frac{\alpha}{\exp(\mu_t)} p_{a,t} y_t, \tag{3.2}$$
$$w_t l_t = (1-\alpha)[r_t k_t + w_t l_t] = \frac{1-\alpha}{\exp(\mu_t)} p_{a,t} y_t. \tag{3.3}$$

Foreign output Y_t^* is produced by foreign intermediate good firms in a symmetric fashion, providing the foreign counterparts of Equations (3.1), (3.2), and ((3.3).

$$y_t^* = \exp(\nu_t^*)(k_t^*)^{\alpha}(l_t^*)^{1-\alpha}, \tag{3.4}$$
$$r_t^* k_t^* = \frac{\alpha}{\exp(\mu_t^*)} p_{a^*,t} y_t^*, \tag{3.5}$$
$$w_t^* l_t^* = \frac{1-\alpha}{\exp(\mu_t^*)} p_{a^*,t} y_t^*. \tag{3.6}$$

Intermediate goods are used as inputs by both home and foreign final goods-producing firms. Particularly, a_t units of the home intermediate output are absorbed locally and x_t units are exported. Analogously, a_t^* units of the foreign intermediate output are used as input by foreign firms, and x_t^* units are exported to the home country.

3.2.2 Final Good Firms

Home and foreign intermediate goods are sold in a frictionless international spot market, from which locally based perfect competitive firms buy inputs to produce home and foreign final consumption goods, g_t and g_t^*, respectively. A typical home final good firm will use local inputs a_t and imported foreign inputs x_t^*, at given prices $[p_{a,t}, p_{x^*,t}]$, which are expressed in local currency. [3] Similarly, a typical foreign final good firm will use local inputs a_t^* and imports x_t, at given prices $[p_{a^*,t}, p_{x,t}]$. Inputs are aggregated in the home and foreign countries through CES technologies:

$$g(a_t, x_t^*) = \left[\omega a_t^{\frac{\sigma-1}{\sigma}} + (1-\omega)(x_t^*)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \tag{3.7}$$

$$g^*(a_t^*, x_t) = \left[\omega^*(a_t^*)^{\frac{\sigma-1}{\sigma}} + (1-\omega^*)x_t^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \tag{3.8}$$

where weights $\omega > 1/2$ and $\omega^* > 1/2$ bias production towards local inputs and thus provide real exchange rate dynamics to the model. Taking the vector of prices $[p_{a,t}, p_{x,t}, p_{a^*,t}, p_{x^*,t}]$ as given, these firms chose local and imported inputs to maximize their profits:

$$\text{Home:} \qquad \max_{\{a_t, x_t^*\}} \quad \{g(a_t, x_t^*) - p_{a,t}a_t - p_{x^*,t}x_t^*\},$$

$$\text{Foreign:} \qquad \max_{\{a_t^*, x_t\}} \quad \{g^*(a_t^*, x_t) - p_{a^*,t}a_t^* - p_{x,t}x_t\}.$$

[3] I use the price of the local final good as *numeraire*. Hence, both locally produced intermediate good prices and imported goods are relative to the local final good price. The real exchange rate s_t will be used to convert foreign currency prices into home currency prices.

The FOC that characterize the solution of these problems are:

$$p_{a,t} = \omega(g_t/a_t)^{1/\sigma}, \tag{3.9}$$

$$p_{x^*,t} = (1-\omega)(g_t/x_t^*)^{1/\sigma}, \tag{3.10}$$

$$p_{a^*,t} = \omega^*(g^*/a_t^*)^{1/\sigma}, \tag{3.11}$$

$$p_{x,t} = (1-\omega^*)(g_t^*/x_t)^{1/\sigma}. \tag{3.12}$$

The real exchange rate s_t is defined as the number of units of home final goods that need to be sold to purchase one unit of foreign final good: [4]

$$s_t = \frac{p_{a,t}}{p_{x,t}} = \frac{p_{x^*,t}}{p_{a^*,t}}. \tag{3.13}$$

Similarly, the terms of trade ratio tt_t is defined as home import prices over home export prices

$$tt_t = \frac{p_{x^*,t}}{p_{a,t}} = s_t \frac{p_{a^*,t}}{p_{a,t}}. \tag{3.14}$$

3.2.3 Capital Managers

Each country's capital is managed by a representative firm owned by the local households. Restricted by a law of motion of capital, these companies choose the optimal sequence

[4] Recall that $p_{x^*,t}$ is the price of foreign inputs expressed in home currency relative to the home final good price, and that $p_{a^*,t}$ is the price of foreign inputs expressed in foreign currency relative to the foreign final good price. Note also that intermediate goods are sold in a frictionless international market, and thus goods produced in the same country should have the same price. If we denote the nominal exchange rate, the home final good price, and the foreign final good price, by E_t, P_t and P_t^*, respectively, Equation (3.13) yields the standard definition of real exchange rate:

$$s_t = \frac{p_{x^*,t}}{p_{a^*,t}} = \frac{(E_t P_{x^*,t})/P_t}{P_{a^*,t}/P_t^*} = E_t \frac{P_t^*}{P_t}.$$

of future capital and investment, $[K_{t+1}, I_t]$, such that capital rents net of investments are maximized.

$$\max_{\{Z_{t+1}k_{t+1}, Z_t ii_t\}} \quad \mathbb{E}_0 \sum_{t=0}^{\infty} \lambda_t [r_t Z_t k_t - Z_t ii_t]$$
$$\text{s.t.} \quad k_{t+1} \exp(g_{z,t+1}) = (1-\delta)k_t + ii_t \qquad \forall t, \tag{3.15}$$

where the term $\exp(g_{z,t+1}) = Z_{t+1}/Z_t$ accounts for the growth rate of the economy, and λ_t stands for local households' weighting of the different periods/states (see subsection 3.2.4). I include the law of motion of capital in the maximization problem to get an unrestricted problem:

$$\max_{\{Z_{t+1}k_{t+1}\}} \quad \mathbb{E}_0 \sum_{t=0}^{\infty} \lambda_t [r_t Z_t k_t - (k_{t+1}Z_{t+1} - (1-\delta)Z_t k_t)].$$

Imposing an interior solution, the FOC gives us

$$1 = \mathbb{E}\big[\frac{\lambda_{t+1}}{\lambda_t}(r_{t+1} + 1 - \delta)\big]. \tag{3.16}$$

Analogously, the equilibrium conditions that characterize the foreign capital management are:

$$k^*_{t+1} \exp(g_{z,t+1}) = (1-\delta)k^*_t + ii^*_t, \tag{3.17}$$
$$1 = \mathbb{E}\big[\frac{\lambda^*_{t+1}}{\lambda^*_t}(r^*_{t+1} + 1 - \delta)\big]. \tag{3.18}$$

Once I have solved the profits of each type of firm, I can define a stock with fixed supply of shares that aggregates the three representative firms for each country and calculate its periodic dividends. Intermediate good firms' profits are given by $p_{a,t}y_t - w_t l_t - r_t k_t$. Capital management firms' profits are $r_t k_t - ii_t$. Finally, final goods firms are perfectly competitive and so their profits are zero. Using Equation (3.3) to replace $w_t l_t$, the dividends of these aggregate country-specific stocks are given by:

$$d_t = p_{a,t} y_t - w_t l_t - ii_t = [1 - (1-\alpha)/\exp(\mu_t)][p_{a,t} y_t] - ii_t, \tag{3.19}$$

$$d_t^* = p_{a^*,t} y_t^* - w_t^* l_t^* - ii_t^* = [1 - (1-\alpha)/\exp(\mu_t^*)][p_{a^*,t} y_t^*] - ii_t^*. \tag{3.20}$$

For each country's stock, I assume that a fraction of shares is owned locally. In particular, home households own a fraction Θ_h of the home stock shares and a fraction $Theta_h^*$ of the foreign stock shares.

3.2.4 Households

A representative household derives utility from consumption C_t and disutility from working hours l_t. Every period, she allocates their resources to consumption, savings, paying lump-sum taxes, and paying consumption transaction costs. Households can save in the international market through home and foreign non-contingent one-period government bonds, $B_{h,t+1}$ and $B_{h,t+1}^*$. Their income can be split between labor income $w_t Z_t l_t$, home and foreign dividends income $\Theta_h D_t + \Theta_h^* s_t D_t^*$, home and foreign bond payments $B_{h,t} + B_{h,t}^* s_t$, and the consumption transaction cost which are rebated back in a lump sump way.

Households choose the optimal sequence of consumption $Z_t c_t$, labor l_t and bond holdings $Z_t b_{h,t+1}$ and $Z_t b_{h,t+1}^*$ that maximize their budget restricted problem.[5]

[5]BKK-type models present a unit root: transitory shocks on productivity produce permanent changes in wealth distribution across countries. Once a country faces a productivity shock, households of that country can save more compared to the other country and that difference does not fade away, producing permanent effects over relative consumption, wealth, etc. One solution, which is used here, is to add (close to zero) bond holding costs. To simplify the exposition these costs are not written.

$$\max_{\{Z_t c_t, l_t, Z_t b_{h,t+1}, Z_t b^*_{h,t+1},\}} \quad \mathbb{E}\sum_{t=0}^{\infty}\beta^t Z_t^{1-\gamma}\left[\frac{c_t^{1-\gamma}}{1-\gamma}-\phi\frac{l_t^{1+1/\epsilon}}{1+1/\epsilon}\right]$$

$$\begin{aligned}\text{s.t.} \quad & Z_t[w_t l_t + \Theta_h d_t + \Theta_h^* s_t d_t^* + \exp(-g_{z,t})(b_{h,t} + b^*_{h,t} s_t) + TC_t] \\ & \qquad \geq Z_t[c_t + q_t b_{h,t+1} + q_t^* b^*_{h,t+1} s_t + \text{tax}_t + \psi_t c_t^{\eta} b_{h,t+1}^{1-\eta}] \qquad \forall t. \end{aligned} \tag{3.21}$$

As it can be guessed, the inclusion of consumption transaction cost is the feature that enables the model to generate endogenous convenience yields. As Krishnamurthy and Vissing-Jorgensen (2012) describe them, convenience assets are securities that provide relatively high liquidity and safety. These features allow these assets to play a role similar to money, easing financial transactions, and thus investors are willing to pay a higher price for them. In this regard, I follow Valchev (2020) to naturally model convenience assets as securities that ease transactions. Particularly, I model consumption transaction costs as a decreasing function of home bond holdings ($\eta > 1$), both for home and foreign households.

The specification of consumption transaction costs is not trivial.[6] Firstly, by only allowing home bonds to ease transactions, I resemble the documented asymmetry in convenience assets seen across countries, where dollar-denominated safe assets hold higher convenience yields (e.g., Krishnamurthy and Vissing-Jorgensen, 2012; Jiang, Krishnamurthy, and Lustig, 2020). This asymmetry will play an important role when addressing cross-country effects of alternative shocks (see subsection 3.3.2). Secondly, I model consumption transaction costs as a decreasing and strictly convex function in home bond holdings. As a consequence, the home bond amount outstanding and the convenience benefits agents extract from them go in opposite directions in equilibrium. This is a feature of convenience assets broadly

[6] In Appendix 3.A.1 I compare our specification with closely related modeling choices.

documented[7]. Finally, to address the time-varying nature of convenience assets demand, I include the term ψ_t, common for home and foreign households, which represents what in the literature is referred to as the "liquidity shock". This process is modeled using the following autoregressive structure:

$$\psi_t = (1 - \rho_\psi)\overline{\psi} + \rho_\psi \psi_{t-1} + u_{\psi,t},$$

where $u_{\psi,t}$ is a zero-mean shock. The FOCs of the household's optimization problem are

$$\begin{aligned}
Z_t c_t : &\qquad \beta^t (Z_t c_t)^{-\gamma} = \lambda_t [1 + \psi_t \eta (c_t / b_{h,t+1})^{\eta - 1}], \\
l_t : &\qquad \beta^t (Z_t)^{1-\gamma} \phi (l_t)^{1/\epsilon} = \lambda_t Z_t w_t, \\
Z_t b_{h,t+1} : &\qquad q_t = \mathbb{E}_t \left[\frac{\lambda_{t+1}}{\lambda_t} \right] - \psi_t (1 - \eta)(c_t / b_{h,t+1})^\eta, \\
Z_t b^*_{h,t+1} : &\qquad q^*_t = \mathbb{E}_t \left[\frac{\lambda_{t+1}}{\lambda_t} \frac{s_{t+1}}{s_t} \right].
\end{aligned}$$

Although stock shares are not traded in the model, let me provide its pricing. As it will be shown in subsection 3.4.1, the inclusion of convenience yields have important consequences for the risk-premium comprised in stocks returns, thus it is convenient to state the equilibrium prices of these assets. For this, denote the home and foreign stock prices in the home country and home currency by $ps_{h,t}$ and $ps^*_{h,t}$, respectively.[8] Manipulating the above FOCs, I get the following optimality conditions:

[7]For recent evidence, see Van Binsbergen, Diamond, and Grotteria (2022)

[8]Each stock will have a home and a foreign price, a feature that can be interpreted as same the same stock being listed in different countries' exchanges.

$$\frac{c_t^{-\gamma}}{1+\psi_t\eta(c_t/b_{h,t+1})^{\eta-1}} = \phi\frac{l_t^{1/\epsilon}}{w_t}, \tag{3.22}$$

$$M_{t,t+1} = \beta\Big[\frac{c_{t+1}}{c_t}\Big]^{-\gamma}\exp(-\gamma g_{z,t+1})\Big[\frac{1+\psi_t\eta(c_t/b_{h,t+1})^{\eta-1}}{1+\psi_{t+1}\eta(c_{t+1}/b_{h,t+2})^{\eta-1}}\Big], \tag{3.23}$$

$$q_t = \mathbb{E}_t\Big[M_{t,t+1}\Big] + \psi_t(\eta-1)(c_t/b_{h,t+1})^{\eta}, \tag{3.24}$$

$$q_t^* = \mathbb{E}_t\Big[M_{t,t+1}\frac{s_{t+1}}{s_t}\Big], \tag{3.25}$$

$$ps_{h,t} = \mathbb{E}_t\Big[M_{t,t+1}\exp(g_{z,t+1})(d_{t+1}+ps_{h,t+1})\Big], \tag{3.26}$$

$$ps_{h,t}^* = \mathbb{E}_t\Big[M_{t,t+1}\exp(g_{z,t+1})(s_{t+1}d_{t+1}^*+ps_{h,t+1}^*)\Big], \tag{3.27}$$

where $M_{t,t+1} = \lambda_{t+1}/\lambda_t$ defines the one period stochastic discount factor. Except for the fact that foreign consumption transaction costs are also decreasing in home bonds, foreign households face a symmetric problem and act optimally following the foreign counterpart of Equations (3.22)-(3.27).

The introduction of consumption transaction costs and convenience assets has an effect both on real activity and on the pricing of financial assets. On the one hand, it increases the marginal cost of consumption, thus households will find it optimal to consume less. This can be seen in Equation (3.22), where, given the same ratio for marginal utility over marginal cost of leisure, households will need a higher marginal utility of consumption to compensate for a higher marginal cost. On the other hand, the introduction of convenience yields affects asset prices in two ways. Firstly, it has a direct effect on the convenience asset price. In Equation (3.24) the standard asset pricing equation for a one-period non-contingent bond is modified, as the expected stochastic discount factor adds to a term that reflects the convenience services: households will be willing to pay a premium for home bonds since its holdings allow them to consume at a lower cost. Secondly, we observe a non-standard multiplicative term in the stochastic discount factor in Equation (3.23). This variable is composed by the growth in the ratio of the marginal utility of consumption and its marginal cost. In most models the price of consumption goods is set to 1 and used as *numeraire*. Hence, the marginal cost of consumption in different periods is just affected by the stochastic growth. When including transaction costs, the marginal cost of consumption is also affected by the ratio of consumption and home bonds that households hold to face that consumption. The evolution of this latter ratio will affect the stochastic discount factor, and hence the price of any asset valuated with such a random variable.

3.2.5 Government

The home and the foreign countries have the same government structure. Here the home government processes and restrictions are described. I assume that the government expenditure Gov_t and debt issuance B_t, as a percentage of GDP, follow exogenous auto-regressive processes:

$$\frac{\text{gov}_t}{y_t p_{a,t}} = (1-\rho_{\text{gov}})\left(\overline{\frac{\text{Gov}}{GDP}}\right) + \rho_{\text{gov}}\frac{\text{gov}_{t-1}}{y_{t-1}p_{a,t-1}} + u_{\text{gov},t}, \tag{3.28}$$

$$\frac{q_t b_t}{y_t p_{a,t}} = (1-\rho_b)\bar{B} + \rho_b\frac{b_{t-1}}{y_{t-1}p_{a,t-1}} + u_{b,t}. \tag{3.29}$$

Governments satisfy their expenditure and their debt burden with lump-sum taxes. [9]

[9] I could alternatively impose more restrictions here, such as a single exogenous process for government expenditures, which are financed partially with debt and partially with taxes:

$$\frac{q_t b_t}{y_t p_{a,t}} = (1-\rho_b)\bar{B} + \rho_b\frac{b_{t-1}}{y_{t-1}p_{a,t-1}} + \kappa u_{\text{gov}},$$

where the parameter κ indicates how much government expenditure is financed with debt and hence how smooth it is the response of taxes to government expenditure shocks.

Their budget constraint is given by

$$q_t b_t = \exp(-g_{z,t}) b_{t-1} + \text{gov}_t - \text{tax}_t. \tag{3.30}$$

To preserve a symmetric steady-state equilibrium with respect to households' allocations in this incomplete market economy, I will impose that both governments collect the same amount of taxes in their local currency. Note that, even though home and foreign governments face the same dynamic equations, the home government will be able to issue more debt in this symmetric equilibrium. Particularly, the fact that home bonds provide convenience services gives them a higher value than foreign bonds. This reduces the relative interest rate paid by the home government and thus allows this government to issue more debt keeping the same debt burden as the foreign one.

3.2.6 Market Clearing

The following conditions guarantee that the intermediate goods market, the final goods market, the labor market, and the government bonds market are cleared.

$$\text{Intermediate Good} \qquad y_t = a_t + x_t, \tag{3.31}$$

$$y_t^* = a_t^* + x_t^*, \tag{3.32}$$

$$\text{Final Good} \qquad g_t = c_t + ii_t + \text{gov}_t, \tag{3.33}$$

$$g_t^* = c_t^* + ii_t^* + \text{gov}_t^*. \tag{3.34}$$

$$\text{Government Debt} \qquad b_t = b_{h,t} + b_{f,t}, \tag{3.35}$$

$$b_t^* = b_{h,t}^* + b_{f,t}^*. \tag{3.36}$$

3.2.7 Exogenous Processes

Aside from the government expenditure and debt issuance processes, the dynamics of the model are driven by 5 additional exogenous processes: home and foreign productivity ν_t and ν_t^*, home and foreign markups μ_t and μ_t^*, and global growth $g_{z,t}$. Markups and productivity shocks have the following structure:

$$\chi_t = (1-\rho_\chi)\chi_t^1 + \rho_\chi \chi_{t-1} + u_{\chi,t},$$
$$\chi_t^1 = (1-\rho_{\chi^1})\overline{\chi} + \rho_{\chi^1}\chi_{t-1}^1 + u_{\chi^1,t},$$

where $u_{\chi,t}$ and $u_{\chi^1,t}$ are the temporary and persistent zero-mean shocks, respectively. In turn, global growth is modeled as

$$g_{z,t} = \overline{g_z} + v_t + u_{g_z,t},$$
$$v_t = \rho_v v_{t-1} + u_{v,t},$$

where $u_{g_z,t}$ and $u_{v,t}$ are zero-mean shocks that represent temporary and persistent deviations from the growth path, respectively.

3.3 Calibration and Model Dynamics

I solve the model using a log-linear second-order approximation around the steady state of our system of equations. This solution method suits the main purpose of this paper, which is to study asset prices' dynamics. Specifically, the approximation includes second moments, and so risk aversion is allowed to affect assets' pricing. In the following paragraphs, I present the parametrization used and how the economy responds when facing alternative shocks. I particularly analyze how the inclusion of convenience yields affects the BKK model dynamics.

3.3.1 Calibration

The model is calibrated at a quarterly frequency. Table 3.1 presents the parameter choices, which are standard in the international macro literature (e.g., Heathcote and Perri, 2002, 2014; Liu, Schmid, and Yaron, 2020). The time discounting parameter β is set at 0.998, the parameter that controls the elasticity labor/wage ϵ is set at 1, and the consumption smoothing parameter γ equals 2. Regarding the parameters that affect firms' decisions, I set a capital share α of 0.36, a depreciation rate δ of 0.015, and an elasticity of substitution σ between local and imported goods of 1.5. Regarding the parameters that govern convenience yields, I follow Valchev (2020) and chose parameters for the transaction costs η and $\overline{\psi}$ to match a steady state convenience yield of around 100 bps, which is in line with the estimations in the literature. Finally, I calibrate the steady-state ratio of government expenditure to GDP at 20% and the ratio of home government debt to GDP at 50%, which is in line with the unconditional 50-year average of the US. Regarding the steady state foreign ratio of government debt to GDP, I solve this in equilibrium and reach a steady state value of 31.6%.[10]

3.3.2 Model Dynamics

Consider first an increase in the debt issuance of the home country. Figures 3.1 and 3.2 show the evolution of a set of selected variables after an increase in home debt b_t, where orange lines depict the case without convenience assets and blue lines represent the case with convenience assets, i.e. with and without consumption transaction costs, respectively.

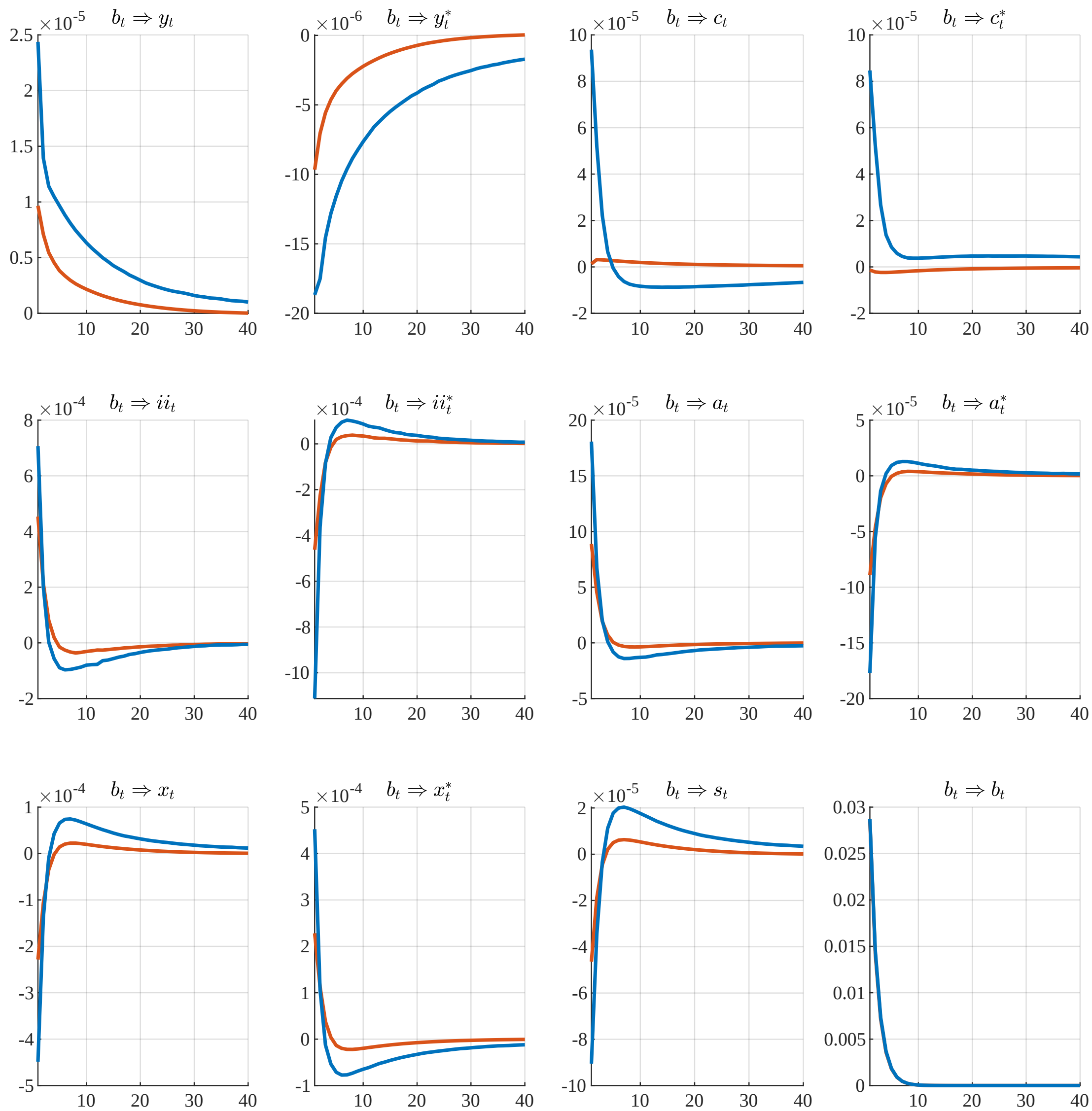
[10] As previously stated, the calibration regarding government processes aims to keep symmetry in the allocations chosen by households. In this regard, government expenditure affects aggregate demand, so it is convenient to use a symmetric steady-state ratio of government expenditure to GDP. Similarly, households' budget constraints are binding in every period so I need to impose the same lump sum taxes for both countries. Since the foreign bond interest rate is higher in equilibrium, the government budget constraint imposes that the steady state foreign debt to GDP ratio is lower than that of the home country.

Table 3.1: Calibrated Parameters. Time Period = 1 quarter.

Parameter	Description	Value
-Households-		
β	Time discount	0.998
ϵ	Elasticity labor/wage	1
γ	Consumption Smoothing	2
η	Transaction cost convexity	2
-Firms-		
α	Capital Share	0.36
δ	Depreciation	0.015
σ	EoS in g,g^*	1.5
Θ_h	Home % of home stocks	0.8
Θ_h^*	Home % of foreign stocks	0.2
-Steady State Targeted Values-		
l	Labor	1/3
tt	Terms of trade	1
x^*/y	Import Share	0.25
Gov/GDP	Gov Expenditure over GDP	0.2
B/GDP	Home Gov Debt over GDP	0.5
B^*/GDP^*	Foreign Gov Debt over GDP	0.32

Let me initially analyze the case where there are no convenience assets ($\psi_t = 0\ \forall t$). In such a case, an increase in the supply of home bonds affects households through two main channels. On the one hand, the decrease in the home bonds price q_t, necessary for such a market to clear, makes both home and foreign households reallocate resources towards larger savings in home bonds, reducing thus their savings in foreign bonds, consumption, and investment. On the other hand, a larger issuance of home bonds relaxes the home government's budget constraint and allows this government to impose smaller taxes on its taxpayers. Home households translate this increase in net income into higher savings, both in home and foreign assets, and into higher consumption and investment.

Figure 3.1: Home Debt IRF - Real Sector.



Note: The graph presents the variables' responses, expressed in percentage deviations from their steady-state values, when the debt-to-GDP ratio suffers a positive shock of one standard deviation of its random component. The variables included are GDP, consumption, investment, inputs absorbed locally, inputs exported, and the real exchange rate, for the home and foreign countries. Orange (blue) lines represent the model without (with) convenience yields. The baseline calibration in Tables 3.1 and3.A.1 is used in both models.

As expected, the effect of the increase in home bond issuance is asymmetric: while home households enjoy both lower bond prices and a reduction in taxes, foreign households are only first-order affected by the reduction in the home bond price. Therefore, the overall change in consumption and investment is only positive for home households. Given that the model features a local bias in the production of final consumption goods, the cross-country asymmetric response in consumption and investment translates into a cross-country asymmetric response in output, where the home output increases and the foreign output decreases. Finally, the different responses of output imply a net inflow of inputs towards the home country and a consequent appreciation of the home currency in real terms.

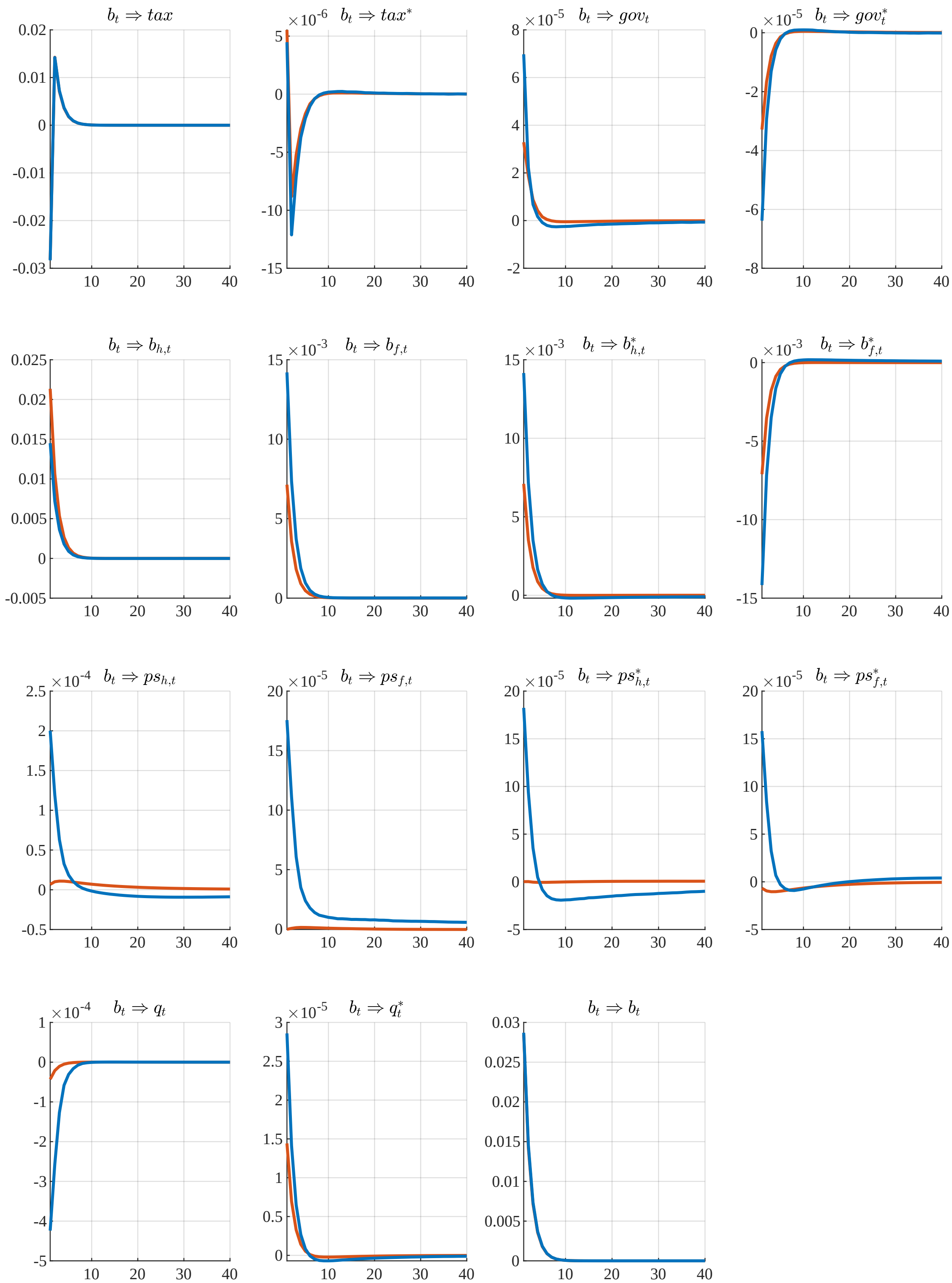
The asymmetric impact of the increase in home debt can also be observed when addressing stock prices. When measured in local currency, home stocks' prices increase and foreign stocks' prices decrease, reflecting the different effects over the stream of dividends each security is expected to pay.[11]

When adding convenience yields, i.e. transaction costs in our model, a third channel is included. The higher supply of convenience assets reduces the equilibrium consumption transaction costs, relaxing both home and foreign households' budget constraints (income effect) and decreasing the effective relative price of consumption (substitution effect). As a consequence, the increase in home bond supply has a larger positive effect on consumption compared to the situation without convenience yields.

Finally, the inclusion of convenience yields makes the home bond price more sensitive to a change in its supply, and thus we observe an amplification of the home bond price channel over the real activity. Since consumption transaction costs are convex, the marginal convenience yield, and in turn the price of each outstanding bond, decreases in the number of home bonds held. Therefore, when convenience yields are considered and there is an increase in the home bond issuance, home bond prices not only decrease because of an increase in the supply but also because of a decrease in the demand, the latter being the consequence

[11]The prices of foreign (home) stocks held by home (foreign) households ps_h^* (ps_f) are not affected, since the exchange rate effect cancels out the dividends effect.

Figure 3.2: Home Debt IRF - Finance Sector.



Note: See note in Figure 3.1. The variables included are taxes, gov expenditure, government bond holdings, prices of stocks, and the price of government bonds, for the home and foreign countries.

of the reduction in the convenience services each bond held provides.

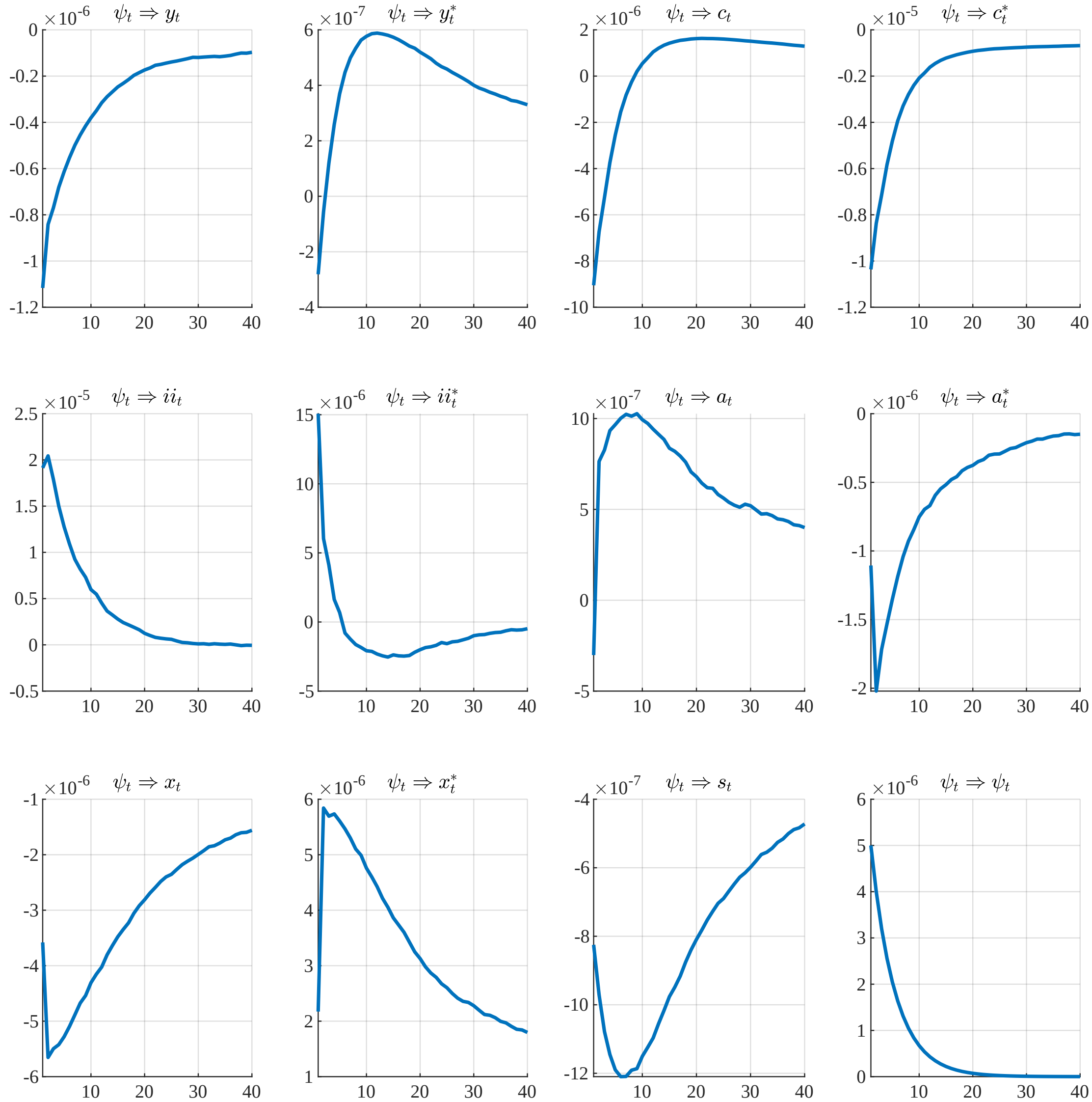
Another way of addressing the role of convenience yields in our model is to analyze a liquidity shock, i.e. an exogenous increase in the consumption transaction costs weight ψ_t. Figures 3.3 and 3.4 show the impulse response functions on such a shock. An exogenous increase in consumption transaction costs has two main effects. On the one hand, it increases the effective cost of consumption in both countries. This produces a substitution effect by which agents reallocate their resources away from consuming final goods and an income effect by which the entire budget constraint shrinks. As a consequence, we observe how consumption is reduced from its steady-state value in both countries. On the other hand, it increases the demand for home bonds, as increasing the holdings of these assets helps households mitigate the negative effect of the liquidity shock. This flight-to-quality movement increases the home bond price q_t, producing a positive revaluation of the home government's debt which allows this government to reduce taxes.[12] As a consequence, home households enjoy an additional net income which translates into more investment and, eight quarters after the liquidity shock takes place, more consumption.

As it can be guessed, the asymmetric response of the local absorption, i.e. consumption and investment, affects the responses of the local output. The better performance of investment and consumption in the home country translates into better performances of the inputs used in such production, a_t and x_t^*, compared to those used in foreign production, a_t and x_t^*. Finally, the positive net flow of inputs entering the home country appreciates the real exchange rate. This evolution of the real exchange rate caused by the liquidity shock is consistent with the patterns depicted in Maggiori (2017).

Note how important asymmetry is in this model. Although both countries face the

[12]Figure 3.4 shows that, after the initial increase due to the lower debt issued, both countries reduce their taxes. This is due to the lower GDP, and thus the lower government expenditure associated with it. We observe that foreign taxes-to-GDP decrease considerably less than that of the home country.

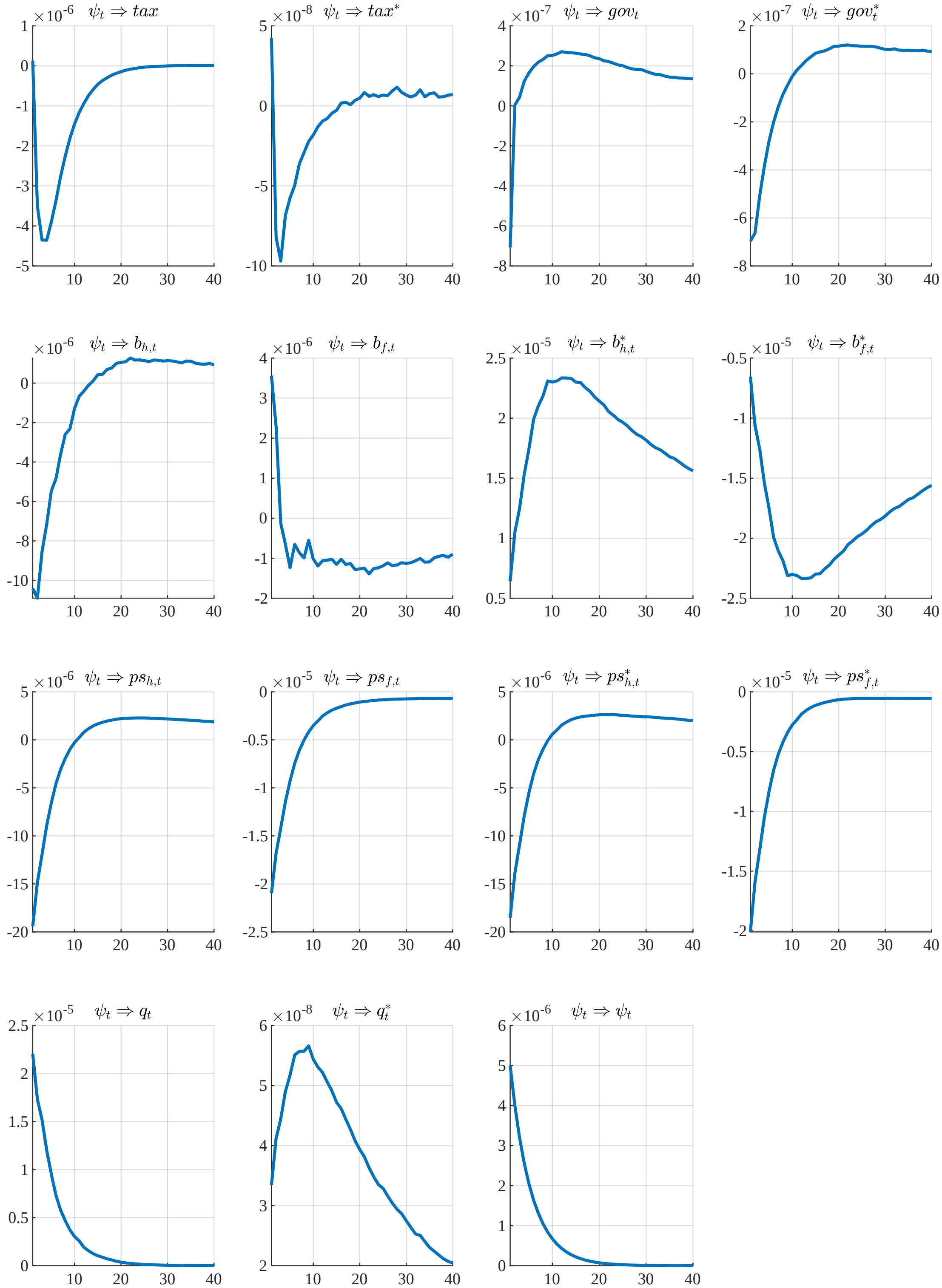
Figure 3.3: Liquidity Shock IRF - Real Sector.



Note: The graph presents the variables' responses, expressed in percentage deviations from their steady-state values, when the transaction costs weight suffers a positive shock of one standard deviation of its random component. The variables included are GDP, consumption, investment, inputs absorbed locally, inputs exported, and the real exchange rate, for the home and foreign countries. The baseline calibration in Tables 3.1 and3.A.1 is used in both models.

same liquidity shock, we see a different effect over both real and financial variables: the convenience asset-issuing country benefits from its higher demand in times of stress, and thus home households have to pay relatively less taxes than their foreign fellows. In other

Figure 3.4: Liquidity Shock IRF - Finance Sector.



Note: See note in Figure 3.3. The variables included are taxes, government expenditure, government bond holdings, prices of stocks, and the price of government bonds, for the home and foreign countries.

words, a liquidity shock in this environment implies a redistribution from the foreign country to the home country: all households pay higher prices to save in convenience assets but only home households benefit from the relaxation in their government budget and the consequent lower taxes.

This mechanism is in sharp contrast with the one proposed in Gourinchas and Rey (2022), where the home country provides insurance to the foreign one. As Maggiori (2017) points out, during times of crises the insurance transfer to the rest of the world would depreciate the home currency. However, these real exchange rate patterns during crisis times appear at odds with the ones observed during 2007-2009. The inclusion of a convenience yield on dollar liabilities can help match the exchange rate dynamics in crisis times.

3.4 Convenience Yields and Asset Pricing Puzzles

In this section, I show that the inclusion of convenience assets can help solve the so-called risk-free, risk premium, and uncovered interest rate parity puzzles. In a nutshell, convenience yields introduce a wedge that increases the risk-free asset price and which varies according to the importance of the liquidity shock and the amount of outstanding debt, replicating thus documented empirical facts that would otherwise be off.

3.4.1 Risk-Free Rate and Risk Premium Puzzle

The risk-free rate (Weil, 1989) and risk premium puzzles (Mehra and Prescott, 1985) refer to the impossibility of a model with time and state separable intertemporal utility function, e.g. CRRA utility, to jointly match the observed low risk-free interest rates and high excess returns of stocks. To better visualize these puzzles and analyze how convenience yields can help solve them, consider a virtual one-period asset with payoff $\tilde{D}_{t+1}$ and return $\tilde{R}_{t+1}$. Under mild assumptions, the model's equilibrium return of such an asset is:[13]

$$\mathbb{E}_t[\tilde{R}_{t+1}] \approx \underbrace{-\ln(\beta) + \gamma \mathbb{E}_t\big[g_{c,t+1}\big] - \frac{1}{2}\gamma^2 \mathrm{Var}_t\big[g_{c,t+1}\big]}_{\text{Risk-free return}} \underbrace{-\frac{1}{2}\mathrm{Var}_t\big[\tilde{d}_{t+1}\big] + \gamma \mathrm{Cov}_t\big[g_{c,t+1}, \tilde{d}_{t+1}\big]}_{\text{Risk premium}},$$

where $g_{c,t+1} = \ln\left(\frac{C_{t+1}}{C_t}\right)$ and $\tilde{d}_{t+1} = \ln\Big(\tilde{D}_{t+1}\Big)$.

The expected return on this virtual asset can be decomposed into the risk-free return and the risk premium. In turn, the drivers underlying these two components are the incentives of agents to smooth consumption across time and across states. Whenever buying the asset helps consumption smoothing, the asset price increases and its expected return decreases, and vice versa.

The risk-free return is explained by three components. First, asset holders need to be compensated to save when future utility is discounted. The second and third components are related to consumption smoothing across time and states, respectively. On the one hand, if consumption is expected to grow, buying a risk-free asset to translate consumption to the future would prevent consumption smoothing, thus the return of such an asset should be higher. On the other hand, consumption smoothing across states makes every certain unit of consumption more valuable when consumption is more volatile, thus the risk-free rate decreases in such volatility.

The risk premium accounts for the extra compensation asset holders should get when the payoffs are state-dependent. In this regard, not only the volatility of the payoff is priced, but also how such payoffs correlate with the consumption available in those states: assets that pay more when consumption is abundant will have a smaller valuation and thus a higher expected return.

The risk-free rate and risk-premium puzzles rely on the fact that the same parameter γ

[13] See Appendix 3.A.3

is used to weigh all these different consumption smoothing drivers simultaneously. Therefore, tensions arise when we try to calibrate γ to empirically match both the risk-free rate and the risk premium. Specifically, we need an extreme utility curvature to account for the observed risk premium and an implausible extreme risk-free rate to offset the implied incentive to smooth consumption across time when agents expect an increasing sequence of consumption. In the model presented, an average risk premium of around 5%, a moderate estimate for the US, requires a value of $\gamma \approx 15$. Aside from being above most of the literature estimates, this value for γ implies an extremely high risk-free rate, of around 25%.

In the following paragraphs, I argue that the inclusion of convenience yields can account for reasonable excess returns without imposing an unreasonable value for the consumption smoothing parameter. The explanation relies on the wedge that convenience services create when pricing the convenience risk-free asset (see Equation (3.24)). In particular, holding the convenience risk-free asset allows households to ease transactions, thus they are willing to forgo a fraction of its return, and the equilibrium risk-free rate is reduced.

For this purpose, I compare the two risk-free rates of the model, i.e. home and foreign bonds' rates, and the risk premiums associated with the different stocks of the model. Recall that only the home bond rate includes a convenience yield, and thus this comparison provides the effects of convenience services over the equilibrium rates of the model.[14].

I provide comparative statics resulting from affecting the two parameters that directly affect convenience yields. Particularly, I start by addressing the case with no convenience yields ($\psi_t = 0\ \forall t$) and show how rates evolve as we increase the convenience services. Later, I analyze how changes in the steady state home bonds-to-GDP ratio affect these figures.

[14] I consider the home bond as the unique provider of convenience services and calibrate the home country with U.S. data. This decision follows, on the one hand, the tradition in the international macroeconomics literature to take the U.S. Treasury rate as the risk-free rate, and on the other hand, the convenience yield literature which provides compelling evidence of the convenience yield attached to U.S. Treasuries (e.g., Krishnamurthy and Vissing-Jorgensen, 2012; Van Binsbergen, Diamond, and Grotteria, 2022).

Table 3.2 shows how the risk-free rate and the risk premium change when we include convenience yields in the model. The first column fixes $\psi_t = 0$, and so expresses the results we would get for our particular calibration in the standard BKK model. The second and third columns incorporate convenience yields, having the third a 50% increase in the transaction costs parameter mean. As can be seen, without the presence of convenience yields the home and foreign bonds provide the same return, a risk-free rate of 4.34%. When we incorporate convenience yields, the risk-free rate goes down significantly, a consequence of the higher price households are willing to pay in order to hold assets that lower the cost of consumption. This effect on the risk-free rate has a direct consequence on the risk premium. Moreover, the change in the risk premium is entirely driven by the change in the risk-free rate, since the returns on stocks remain constant under different values of $\overline{\psi}$.

Table 3.2: Risk-Free Rate and Risk Premium. Sensitivity on $\overline{\psi}$.

	$\psi_t = 0$	$\overline{\psi} = 0.005$	$\overline{\psi} = 0.0075$
i_t	4.34%	3.25%	2.74%
	(0.016)	(0.017)	(0.018)
i_t^*	4.34%	4.34%	4.33%
	(0.016)	(0.016)	(0.016)
$i_{h,t}^s - i_t$	0.14%	1.23%	1.73%
	(0.002)	(0.005)	(0.006)
$i_{h,t}^{s,*} - i_t$	0.2%	1.31%	1.81%
	(0.002)	(0.005)	(0.006)

Note: i_t and i_t^* are the annualized returns of the home and foreign bonds, respectively; $i_{h,t}^s$ and $i_{h,t}^{s,*}$ are the annualized returns of the home and foreign stocks, respectively, priced in the home market. See Appendix 3.A.4 for the detailed construction of these returns. Moments presented are the mean and the standard deviation (in parenthesis).

Although affecting $\overline{\psi}$ is the most straightforward way of addressing the effects of convenience assets, the reduced form nature of this parameter complicates its mapping to observables, thus preventing the making of testable implications. To overcome this problem, I next study how changes in the average home debt-to-GDP ratio affect the risk-free rate and thus the risk premium.

Table 3.3 shows that the implications of decreasing the debt-to-GDP ratio are similar to those of increasing the transaction costs parameter. The scarcity of bonds increases the transaction costs of each unit of consumption. Consequently, the demand for convenience assets to face those higher costs increases and the convenience risk-free rate decreases. Moreover, since transaction costs are convex ($\eta > 1$), the effect is non-linear: reducing the debt-to-GDP ratio 20 p.p from the baseline calibration reduces the risk-free rate in 200 bps while increasing that ratio by the same amount increases the risk-free rate in 49 bps. This non-linearity is of main importance if we take into account that the main issuer of convenience assets, US, had increased its debt-to-GDP ratio almost constantly since the 80s, with a big jump after the 2008 financial crisis.

Table 3.3: Risk-Free Rate and Risk Premium. Sensitivity on $\overline{B/GDP}$.

	$\overline{B/GDP} =$ 30%	$\overline{B/GDP} =$ 50%	$\overline{B/GDP} =$ 70%
i_t	1.25%	3.25%	3.74%
	(0.022)	(0.017)	(0.016)
i_t^*	4.33%	4.34%	4.34%
	(0.016)	(0.016)	(0.016)
$i_{h,t}^s - i_t$	3.22%	1.23%	0.73%
	(0.014)	(0.005)	(0.003)
$i_{h,t}^{s,*} - i_t$	3.3%	1.31%	0.81%
	(0.014)	(0.005)	(0.003)

Note: i_t and i_t^* are the annualized returns of the home and foreign bonds, respectively; $i_{h,t}^s$ and $i_{h,t}^{s,*}$ are the annualized returns of the home and foreign stocks, respectively, priced in the home market. See Appendix 3.A.4 for the detailed construction of these returns. Moments presented are the mean and the standard deviation (in parenthesis).

The dynamics of convenience yields presented here are in line with the estimates obtained in Van Binsbergen, Diamond, and Grotteria (2022). The authors study the effects of the alternative rounds of Quantitative Easing (QE) on convenience yields. Using small time windows around the dates and times at which announcements were made, the authors estimate that Q.E. 1 reduced 12 months convenience yields by 42 bps, while Q.E 2 and Q.E 3 had no clear effect. That is, the impact of an increase in the supply of convenience assets was higher when their relative supply was lower.

3.4.2 UIP Puzzle

The uncovered interest rate parity (UIP) is an arbitrage condition that relates the home and foreign interest rates with the expected evolution of the exchange rate. Basically, any difference between home and foreign interest rates should be explained by the expected evolution of the exchange rate and the compensation for exchange rate fluctuations, i.e. the foreign exchange risk premium. Although how intuitive and predominant the UIP condition is in international-macro models, there is substantial evidence about persistent deviations from it (Fama, 1984; Engel, 2014). Moreover, these deviations hold when removing the foreign exchange risk premium through exchange rate forward contracts (Du, Tepper, and Verdelhan, 2018b).

To understand this puzzle in the context of our model and how the introduction of convenience yields may help to reconcile the theory with the evidence, I manipulate Equations (3.24) and (3.25) to get the modified UIP condition:[15]

$$i_t - i_t^* = \mathbb{E}_t[g_{s,t+1}] + \frac{1}{2}\mathrm{Var}_t[g_{s,t+1}] + \mathrm{Cov}_t[m_{t,t+1}, g_{s,t+1}] - i_t^c, \tag{3.37}$$

where i_t and i_t^* denote the home and foreign interest rates, $g_{s,t+1}$ is the real depreciation rate of the home currency, $m_{t,t+1}$ is the log of the home stochastic discount factor, and i_t^c is the convenience yield, defined as the rate that a household is willing to forgo to hold the convenience asset.

[15] See Appendix 3.A.3 for the derivation

As can be seen, in the modified UIP condition home interest rates are allowed to be smaller than foreign ones, even when controlling for exchange rate fluctuations. The higher the convenience services home bonds provide, the higher this deviation is in equilibrium. In other words, as home bonds allow to reduce transaction costs, households require a smaller home bond return to be indifferent between investing locally or abroad.

The inclusion of a financial wedge is a natural way to replicate UIP deviations in BKK-type models (e.g., Itskhoki and Mukhin, 2021), where the exchange rate is completely defined in the trade sector.[16] In this environment, households invest in home and foreign risk-free bonds taking the exchange rate evolution as given, thus UIP holds unless some wedge is introduced in the financial markets. I argue that convenience yields can play that role.

Table 3.4 presents the difference in home and foreign interest rates for alternative scenarios of transaction costs. Whenever transaction costs are null, home interest rates are the same as foreign interest rates. This equality holds despite exchange rate fluctuations, as the exchange rate depreciation has zero mean and negligible variance in equilibrium. As the transaction costs parameter $\overline{\psi}$ increases, UIP deviations become larger, as suggested in Equation (3.37).

The inclusion of convenience yields to our international macro model accommodates not only UIP deviations but also the fact that these deviations fade away as we consider longer-term bonds (Chinn, 2006; Du, Im, and Schreger, 2018a; Van Binsbergen, Diamond, and Grotteria, 2022). To see why, denote by q_t^T and $q_t^{*,T}$ the shadow prices of bonds with time-to-maturity of T periods issued by the home and foreign country, respectively, each one

[16] Manipulating the optimality conditions of the final goods firms, and adding the definition of the real exchange rate, we can obtain that the latter is a function of the ratio of imported/local inputs:

$$s_t = \frac{\omega\left[\omega + (1-\omega)\left(\frac{x_t^*}{a_t}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}}}{(1-\omega^*)\left[\omega^*\left[\frac{\omega}{1-\omega}\frac{\omega^*}{1-\omega^*}\right]^{\frac{\sigma-1}{1}}\left(\frac{x_t^*}{a_t}\right)^{\frac{\sigma-1}{\sigma}} + (1-\omega^*)\right]^{\frac{1}{\sigma-1}}}.$$

Table 3.4: UIP deviations. Sensitivity on $\overline{\psi}$.

	$\psi_t = 0$	$\overline{\psi} = 0.005$	$\overline{\psi} = 0.0075$
i_t	4.34%	3.25%	2.74%
	(0.016)	(0.017)	(0.018)
$i_t - i_t^*$	0%	−1.09%	−1.59%
	(0.011)	(0.011)	(0.012)
i_t^∞	4.49%	4.48%	4.47%
	(0.016)	(0.016)	(0.016)
$i_t^\infty - i_t^{*,\infty}$	0.01%	−0%	−0.01%
	(0.011)	(0.010)	(0.010)

Note: i_t and i_t^∞ are the interest rates on one period and consol home bonds, respectively, and i_t^* and $i^{*,\infty}$ are the interest rates on one period and consol foreign bonds, respectively. See Appendix 3.A.4 for the detailed construction of these returns. All rates are annualized. Moments presented are the mean and the standard deviation (in parenthesis).

paying one unit of consumption of the corresponding country per period. These prices equal:

$$q_t^T = \mathbb{E}_t\Big[M_{t,t+1}(1 + q_{t+1}^{T-1})\Big] + \psi_t(\eta - 1)(c_t/B_{h,t+1})^\eta, \tag{3.38}$$

$$q_t^{*,T} = \mathbb{E}_t\Big[M_{t,t+1}(1 + q_{t+1}^{*,T-1})\Big]. \tag{3.39}$$

In contrast with the price of the one-period home bond, the long-term home bond price embeds the convenience premium into a recursive structure, thus it is affected by the entire sequence of expected convenience services:

$$q_t^T = \mathbb{E}_t\Big[\sum_{j=0}^{T-1} M_{t,t+1+j}\Big] + \mathbb{E}_t\Big[\sum_{j=0}^{T-1} M_{t,t+j}\psi_{t+j}(\eta - 1)(c_{t+j}/B_{h,t+1+j})^\eta\Big],$$

$$\text{where} \quad M_{t,t+j} = \Pi_{s=1}^{j} M_{t+s-1,t+s} = \beta^j \Big[\frac{c_{t+j}Z_{t+j}}{c_t Z_t}\Big]^{-\gamma} \Big[\frac{1 + \psi_t\eta(c_t/b_{h,t+1})^{\eta-1}}{1 + \psi_{t+j}\eta(c_{t+j}/B_{h,t+j+1})^{\eta-1}}\Big].$$

This recursive structure implies that the price of the bond today shares a common sequence of convenience services with the price of the bond in the future. Therefore, when computing returns, the premiums in the denominator (price of the bond today) and numerator (price of the bond tomorrow plus periodic payments) tend to cancel out as the time to maturity increases. Table 3.4 exemplifies these findings for the extreme case, i.e. for consol bonds. Whenever transaction costs' parameter increase we observe a deviation in UIP in the short run but not in the long horizon.

3.5 Conclusion

In this paper, I extended the two-country model in Backus, Kehoe, and Kydland (1992) by incorporating endogenous convenience yields. I do so by including a transaction cost technology which is decreasing in the amount of home bonds held by households. This extension generates an asymmetry in an otherwise symmetric model: the country that issues convenience assets can roll over its debt at a lower cost and thus needs to collect lower taxes from its households. I study two different shocks. I find that the new issuance of convenience assets spills over to foreign households, as their equilibrium transaction costs are reduced. I also show that a global liquidity shock affects both countries differently, as the pricing of convenience assets increases in this shock and thus allows the issuing country to reduce taxes. Finally, I show that the inclusion of convenience yields helps the model to reconcile empirical patterns of rates of returns across assets and countries with its theoretical predictions. In particular, the model is able to obtain plausible equilibrium risk premiums and risk-free rates without imposing an extreme risk aversion. Moreover, consistent with the empirical literature, the model generates UIP deviations that fade away when considering long-term bonds.

3.A Appendix

3.A.1 Convenience Yields Alternative Modelling Choices

The model presented introduces consumption transaction costs as a device to generate endogenous convenience yields. This reduced-form approach to model convenience yields is related to models that obtain a positive value for currency through the introduction of shopping time[17]. In those models, consumption takes time and the time spent is a (transaction cost) function that depends negatively on the amount of money held by the households. Since shopping and leisure time add up to a fixed constant, transaction costs are weighted by the marginal utility of leisure in each period. In our specification, on the contrary, transaction cost will be weighted by the marginal utility of consumption in each period, and so our specification will yield different dynamics than that of shopping time models.

To see this more clearly, here I explicitly derive the optimality conditions that I would obtain if I were to use the latter approach. Agents would distribute their time between leisure le_t, labor l_t, and shopping s_t, facing an additional restriction $1 - le_t = s_t + l_t$. Using the same specification as in the presented model, $s_t = \psi_t c_t^{\eta} B_{h,t+1}^{1-\eta}$, computing first order conditions and rearranging, we get:

$$\begin{aligned}
Z_t c_t: &\qquad \beta^t (Z_t c_t)^{-\gamma} = \lambda_t + \zeta_t[\psi_t \eta (c_t/b_{h,t+1})^{\eta-1}] \\
l_t: &\qquad \beta^t (Z_t)^{1-\gamma} \phi(l_t)^{1/\epsilon} + \zeta_t = \lambda_t Z_t w_t \\
Z_t b_{h,t+1}: &\qquad q_t = \mathbb{E}_t\Big[\frac{\lambda_{t+1}}{\lambda_t}\Big] - \frac{\zeta_t}{\lambda_t}\psi_t(1-\eta)(c_t/b_{h,t+1})^{\eta}
\end{aligned}$$

where ζ_t is the Lagrange multiplier associated to our new restriction. The remanding FOCs are equal to our main model, though the λ_t here has a different solution. If $\zeta_t = \lambda_t$, then

[17] For an introduction to shopping time, see Ljungqvist and Sargent (2018), Chapter 24.

transaction costs and shopping time approaches yield a similar solution.

Finally, another alternative to motivate convenience yields is to include bonds in the utility, as in Dekle, Jeong, and Kiyotaki (2013). Feenstra (1986) shows that, when modeling fiat currency, transaction costs and money (bonds) in the utility yield similar results.

3.A.2 Calibration

Table 3.A.1: Calibrated Exogenous Processes Parameters. Time Period = 1 quarter.

Parameter	Description	Value
$\overline{g_z}$	Global growth: mean	0.0045
$\sigma_{u_{g_z}}$	Global growth: std dev temp. shock	0.01
ρ_v	Global growth: persistence perm. shock	0.95
σ_{u_v}	Global growth: std dev perm. shock	0.001
$\overline{\mu}$	Markup: mean	0.05
ρ_μ	Markup: persistence temp. shock	0.95
σ_{u_μ}	Markup: std. dev temp. shock	0.01
ρ_{μ_1}	Markup: persistence perm. shock	0.95
$\sigma_{u_{\mu_1}}$	Markup: std. dev. perm. shock	0.01
$\overline{\nu}$	Productivity: mean	0
ρ_ν	Productivity: persistence temp. shock	0.97
σ_{u_ν}	Productivity: std. dev. temp. shock	0.02
ρ_{ν_1}	Productivity: persistence perm. shock	0.9
$\sigma_{u_{\nu_1}}$	Productivity: std. dev. perm. shock	0.1
ρ_{gov}	Gov. expenditure shock persistence	0.5
$\sigma_{u_{\text{gov}}}$	Gov. expenditure shock std. dev.	0.05
ρ_b	Gov. debt shock persistence	0.5
σ_{u_b}	Gov. debt shock std. dev.	0.05
$\overline{\psi}$	Transaction costs: mean	$5e-4$
ρ_ψ	Transaction costs: persistence shock	0.8
σ_{u_ψ}	Transaction costs: std. dev. shock	$5e-6$

3.A.3 Asset Pricing Puzzles Equations.

I initially address the Risk-free Rate and Risk Premium puzzles. Consider a virtual one-period asset with price $\tilde{P}_t$, next period pay-off $\tilde{D}_{t+1}$ and return $\tilde{R}_{t+1}$. If the random variable formed by the stochastic discount factor times the payoff is log-normal distributed, the equilibrium price of such an asset is:

$$
\begin{aligned}
\tilde{P}_t &= \mathbb{E}_t[M_{t,t+1}\tilde{D}_{t+1}] \\
&= \mathbb{E}_t\Big[\beta\Big(\frac{C_{t+1}}{C_t}\Big)^{-\gamma}\tilde{D}_{t+1}\Big] \\
&= \beta\mathbb{E}_t\Big[\exp\Big(\ln\Big(\Big(\frac{C_{t+1}}{C_t}\Big)^{-\gamma}\tilde{D}_{t+1}\Big)\Big)\Big] \\
&= \beta\mathbb{E}_t\Big[\exp\Big(-\gamma g_{c,t+1} + \tilde{d}_{t+1}\Big)\Big] \\
&= \beta\exp\Big[-\gamma\mathbb{E}_t\big[g_{c,t+1}\big] + \mathbb{E}_t\big[\tilde{d}_{t+1}\big] + \frac{1}{2}\gamma^2\mathrm{Var}_t\big[g_{c,t+1}\big] + \frac{1}{2}\mathrm{Var}_t\big[\tilde{d}_{t+1}\big] - \gamma\mathrm{Cov}_t\big[g_{c,t+1}, \tilde{d}_{t+1}\big]\Big],
\end{aligned}
$$

where $g_{c,t+1} = \ln\left(\frac{C_{t+1}}{C_t}\right)$ and $\tilde{d}_{t+1} = \ln\Big(\tilde{D}_{t+1}\Big)$. The one-period return of such an asset is:

$$
\begin{aligned}
\mathbb{E}_t[\tilde{R}_{t+1}] &\approx \mathbb{E}_t\big[\ln\Big(\tilde{D}_{t+1}/\tilde{P}_t\Big)\big] \\
&= \mathbb{E}_t\big[\tilde{d}_{t+1} - \ln\Big(\tilde{P}_t\Big)\big] \\
&= -\ln(\beta) + \gamma\mathbb{E}_t\big[g_{c,t+1}\big] - \frac{1}{2}\gamma^2\mathrm{Var}_t\big[g_{c,t+1}\big] - \frac{1}{2}\mathrm{Var}_t\big[\tilde{d}_{t+1}\big] + \gamma\mathrm{Cov}_t\big[g_{c,t+1}, \tilde{d}_{t+1}\big]
\end{aligned}
$$

In turn, consider now an asset with a one-time risk-free payoff of 1. Its price and return are denoted by $\tilde{P}^{rf}_t$ and $\tilde{R}^{rf}_{t+1}$ respectively:

$$
\begin{aligned}
\mathbb{E}_t[\tilde{R}^{rf}_{t+1}] &\approx \mathbb{E}_t\big[\ln\Big(1/\tilde{P}^{rf}_t\Big)\big] \\
&= -\ln(\beta) + \gamma\mathbb{E}_t\big[g_{c,t+1}\big] - \frac{1}{2}\gamma^2\mathrm{Var}_t\big[g_{c,t+1}\big]
\end{aligned}
$$

The Risk Premium is defined as the excess return on the asset compared to that of the risk-free asset:

$$
\begin{aligned}
\text{Risk Premium} &\equiv \mathbb{E}_t[\tilde{R}_{t+1} - \tilde{R}^{rf}_{t+1}] \\
&\approx -\frac{1}{2}\text{Var}_t\big[\tilde{d}_{t+1}\big] + \gamma\text{Cov}_t\big[g_{c,t+1}, \tilde{d}_{t+1}\big]
\end{aligned}
$$

The UIP condition relates the home and foreign exchange rates through the expected evolution of the exchange rate. To derive this relation, I start by computing the foreign interest rate. From Equation (3.25):

$$
\begin{aligned}
q^*_t &= \mathbb{E}_t\big[M_{t,t+1}\frac{s_{t+1}}{s_t}\big], \\
&= \mathbb{E}_t\big[\exp(m_{t,t+1} + g_{s,t+1})\big], \\
&= \exp\big(\mathbb{E}_t[m_{t,t+1}] + \mathbb{E}_t[g_{s,t+1}] + \frac{1}{2}\text{Var}_t[m_{t,t+1}] + \frac{1}{2}\text{Var}_t[g_{s,t+1}] + \text{Cov}_t[m_{t,t+1}, g_{s,t+1}]\big), \\
i^*_t &\approx -\ln(q^*_t) \\
&= -\big(\mathbb{E}_t[m_{t,t+1}] + \mathbb{E}_t[g_{s,t+1}] + \frac{1}{2}\text{Var}_t[m_{t,t+1}] + \frac{1}{2}\text{Var}_t[g_{s,t+1}] + \text{Cov}_t[m_{t,t+1}, g_{s,t+1}]\big)
\end{aligned}
$$

where $m_{t,t+1}$ is the log of the home stochastic discount factor and $g_{s,t+1}$ is the real depreciation rate of the home currency.[18].

Next, I compute the home interest rate i_t. For this, define i^c_t as the convenience yield, i.e. the rate that a household is willing to forgo to hold the convenience asset. Since the term $\psi_t(\eta-1)(c_t/B_{h,t+1})^\eta$ is the excess price that households pay to hold home bonds, the

[18]To get the third step I use the log normality of the joint random variable $\big[m_{t,t+1}g_{s,t+1}\big]$.

return of a virtual non-convenient home bond is given by:

$$
\begin{aligned}
i_t + i_t^c &\approx -\ln\left(q_t - \psi_t(\eta - 1)(c_t/B_{h,t+1})^{\eta}\right), \\
&= -\ln\left(\mathbb{E}_t[M_{t,t+1}]\ln\right), \\
&= -\mathbb{E}_t[m_{t+1}] - \frac{1}{2}\text{Var}_t[m_{t+1}],
\end{aligned}
$$

where I use Equation (3.24) in the second step. Finally, compute $i_t + i_t^c - i_t^*$ and rearrange to obtain the expression that represents deviations from the UIP as a function of the convenience yield:

$$
i_t - i_t^* = \mathbb{E}_t[g_{s,t+1}] + \frac{1}{2}\text{Var}_t[g_{s,t+1}] + \text{Cov}_t[m_{t,t+1}, g_{s,t+1}] - i_t^c.
$$

3.A.4 Interest rates computations.

The annualized interest rates presented in Table 3.2, Table 3.3, and Table 3.4 are computed in the following way:

$$
\begin{aligned}
i_t &= (1/q_t)^4 - 1, \\
i_t^* &= (1/q_t^*)^4 - 1, \\
i_{h,t}^s &= \left[g_{z,t+1} * (ps_{h,t+1} + d_{t+1})/ps_{h,t}\right]^4 - 1, \\
i_{h,t}^{s,*} &= \left[g_{z,t+1} * (ps_{h,t+1}^* + s_{t+1}d_{t+1}^*)/ps_{h,t}^*\right]^4 - 1, \\
i_t^{\infty} &= ((1 + q_{t+1}^{\infty})/q_t^{\infty})^4 - 1, \\
i_t^{*,\infty} &= ((1 + q_{t+1}^{*,\infty})/q_t^{*,\infty})^4 - 1.
\end{aligned}
$$

Bibliography

Acharya, V. V., and Pedersen, L. H. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics*, *77*(2), 375–410.

Adrian, T., Boyarchenko, N., and Shachar, O. (2017). Dealer balance sheets and bond liquidity provision. *Journal of Monetary Economics*, *89*, 92–109.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56.

An, Y. (2022). Competing with inventory in dealership markets. *Available at SSRN 3284836*.

An, Y., and Zheng, Z. (2023). Immediacy provision and matchmaking. *Management Science*, *69*(2), 1245–1263.

Anderson, M., and Stulz, R. M. (2017). Is post-crisis bond liquidity lower? Tech. rep., National Bureau of Economic Research.

Backus, D. K., Kehoe, P. J., and Kydland, F. E. (1992). International real business cycles. *Journal of Political Economy*, *100*(4), 745–775.

Bansal, R., and Coleman, W. J. (1996). A monetary explanation of the equity premium, term premium, and risk-free rate puzzles. *Journal of Political Economy*, *104*(6), 1135–1171.

Bao, J., O'Hara, M., and Zhou, X. A. (2018). The volcker rule and corporate bond market making in times of stress. *Journal of Financial Economics*, *130*(1), 95–113.

Bessembinder, H., Jacobsen, S., Maxwell, W., and Venkataraman, K. (2018). Capital commitment and illiquidity in corporate bonds. *The Journal of Finance*, *73*(4), 1615–1661.

Brugler, J., Comerton-Forde, C., and Martin, J. S. (2022). Secondary market transparency and corporate bond issuing costs. *Review of Finance*, *26*(1), 43–77.

Chinn, M. D. (2006). The (partial) rehabilitation of interest rate parity in the floating rate era: Longer horizons, alternative expectations, and emerging markets. *Journal of International Money and Finance*, *25*(1), 7–21.

Choi, J., Huh, Y., and Seunghun Shin, S. (2024). Customer liquidity provision: Implications for corporate bond transaction costs. *Management Science*, *70*(1), 187–206.

Cimon, D., and Garriott, C. (2019). Banking regulation and market making. *Journal of Banking & Finance*, *109*, 105653.

Coen, J., and Coen, P. (2022). A structural model of liquidity in over-the-counter markets.

Cohen, A., Kargar, M., Lester, B., and Weill, P.-O. (2022). Inventory, market making, and liquidity: Theory and application to the corporate bond market.

Colliard, J.-E., Foucault, T., and Hoffmann, P. (2021). Inventory management, dealers' connections, and prices in over-the-counter markets. *The Journal of Finance*, *76*(5), 2199–2247.

Dekle, R., Jeong, H., and Kiyotaki, N. (2013). Dynamics of firms and trade in general equilibrium. *KDI School of Pub Policy & Management Paper*, (13-06).

Di Maggio, M., Franzoni, F., Kermani, A., and Sommavilla, C. (2019). The relevance of broker networks for information diffusion in the stock market. *Journal of Financial Economics*, *134*(2), 419–446.

Dick-Nielsen, J., and Poulsen, T. K. (2019). How to clean academic trace data. *Available at SSRN 3456082*.

Dick-Nielsen, J., and Rossi, M. (2019). The cost of immediacy for corporate bonds. *The Review of Financial Studies*, *32*(1), 1–41.

Du, W., Im, J., and Schreger, J. (2018a). The us treasury premium. *Journal of International Economics*, *112*, 167–181.

Du, W., Tepper, A., and Verdelhan, A. (2018b). Deviations from covered interest rate parity. *The Journal of Finance*, *73*(3), 915–957.

Duffie, D. (2012). Market making under the proposed volcker rule. *Rock Center for Corporate Governance at Stanford University Working Paper*, (106).

Duffie, D. (2017). *Post-crisis bank regulations and financial market liquidity*. Banca d'Italia.

Duffie, D., Fleming, M. J., Keane, F. M., Nelson, C., Shachar, O., and Van Tassel, P. (2023). Dealer capacity and us treasury market functionality. *FRB of New York Staff Report*, (1070).

Duffie, D., Gârleanu, N., and Pedersen, L. H. (2005). Over-the-counter markets. *Econometrica*, *73*(6), 1815–1847.

Duffie, D., Gârleanu, N., and Pedersen, L. H. (2007). Valuation in over-the-counter markets. *The Review of Financial Studies*, *20*(6), 1865–1900.

Dyskant, L., Silva, A. F., and Sultanum, B. (2023). Dealer costs and customer choice. Mimeo.

Edwards, A. K., Harris, L. E., and Piwowar, M. S. (2007). Corporate bond market transaction costs and transparency. *The Journal of Finance*, *62*(3), 1421–1451.

Engel, C. (2014). Exchange rates and interest parity. *Handbook of International economics*, *4*, 453–522.

Fama, E. F. (1984). Forward and spot exchange rates. *Journal of Monetary Economics*, *14*(3), 319–338.

Feenstra, R. C. (1986). Functional equivalence between liquidity costs and the utility of money. *Journal of Monetary Economics*, *17*(2), 271–291.

Feldhütter, P. (2012). The same bond at different prices: identifying search frictions and selling pressures. *The Review of Financial Studies*, *25*(4), 1155–1206.

Foucault, T., Pagano, M., and Röell, A. (2013). *Market liquidity: theory, evidence, and policy*. Oxford University Press, USA.

Friewald, N., and Nagler, F. (2019). Over-the-counter market frictions and yield spread changes. *The Journal of Finance*, *74*(6), 3217–3257.

Goldstein, M. A., and Hotchkiss, E. S. (2020). Providing liquidity in an illiquid market: Dealer behavior in us corporate bonds. *Journal of Financial Economics*, *135*(1), 16–40.

Gorton, G. B. (2010). *Slapped by the invisible hand: The panic of 2007*. Oxford University Press.

Gourinchas, P.-O., and Rey, H. (2022). Exorbitant privilege and exorbitant duty.

Greenwood, R., Stein, J. C., Hanson, S. G., and Sunderam, A. (2017). Strengthening and streamlining bank capital regulation. *Brookings Papers on Economic Activity*, *2017*(2), 479–565.

Greenwood, R., and Vayanos, D. (2014). Bond supply and excess bond returns. *The Review of Financial Studies*, *27*(3), 663–713.

Heathcote, J., and Perri, F. (2002). Financial autarky and international business cycles. *Journal of Monetary Economics*, *49*(3), 601–627.

Heathcote, J., and Perri, F. (2014). Assessing international efficiency. In *Handbook of International Economics*, vol. 4, (pp. 523–584). Elsevier.

Hendershott, T., Li, D., Livdan, D., and Schürhoff, N. (2020). True cost of immediacy. *Swiss Finance Institute Research Paper*, (20-71).

Hendershott, T., Livdan, D., and Schürhoff, N. (2021). All-to-all liquidity in corporate bonds. *Swiss Finance Institute Research Paper*, (21-43).

Hendershott, T., and Madhavan, A. (2015). Click or call? auction versus search in the over-the-counter market. *The Journal of Finance*, *70*(1), 419–447.

Hugonnier, J. (2012). Speculative behavior in decentralized markets.

Hugonnier, J., Lester, B., and Weill, P.-O. (2020). Frictional intermediation in over-the-counter markets. *The Review of Economic Studies*, *87*(3), 1432–1469.

Itskhoki, O., and Mukhin, D. (2021). Exchange rate disconnect in general equilibrium. *Journal of Political Economy*, *129*(8), 2183–2232.

Jiang, Z., Krishnamurthy, A., and Lustig, H. (2020). Dollar safety and the global financial cycle. Tech. rep., National Bureau of Economic Research.

Kargar, M., Lester, B., Lindsay, D., Liu, S., Weill, P.-O., and Zúñiga, D. (2021). Corporate bond liquidity during the covid-19 crisis. *The Review of Financial Studies*, *34*(11), 5352–5401.

Kirkby, R. (2017). Convergence of discretized value function iteration. *Computational Economics*, *49*(1), 117–153.

Krishnamurthy, A., and Vissing-Jorgensen, A. (2012). The aggregate demand for treasury debt. *Journal of Political Economy*, *120*(2), 233–267.

Lagos, R., and Rocheteau, G. (2009). Liquidity in asset markets with search frictions. *Econometrica*, *77*(2), 403–426.

Lagos, R., Rocheteau, G., and Wright, R. (2017). Liquidity: A new monetarist perspective. *Journal of Economic Literature*, *55*(2), 371–440.

Li, J. S., O'Hara, M., Rapp, A. C., and Zhou, X. A. (2023). Bond market illiquidity: Is portfolio trading the solution? *SMU Cox School of Business Research Paper No. 23-09, Available at SSRN 4495516*.

Liu, Y., Schmid, L., and Yaron, A. (2020). The risks of safe assets. *Available at SSRN 3699618*.

Ljungqvist, L., and Sargent, T. J. (2018). *Recursive macroeconomic theory*.

Longstaff, F. A. (2004). The flight-to-liquidity premium in us treasury bond prices. *The Journal of Business*, *77*(3), 511–526.

Maggiori, M. (2017). Financial intermediation, international risk sharing, and reserve currencies. *American Economic Review*, *107*(10), 3038–3071.

Markowits, H. M. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 71–91.

Mehra, R., and Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of monetary Economics*, *15*(2), 145–161.

Meli, J., and Todorova, Z. (2022). Portfolio trading in corporate bond markets. *Available at SSRN 4295453*.

Miao, J. (2006). A search model of centralized and decentralized trade. *Review of Economic dynamics*, *9*(1), 68–92.

Nagel, S. (2016). The liquidity premium of near-money assets. *The Quarterly Journal of Economics*, *131*(4), 1927–1971.

O'Hara, M., and Zhou, X. A. (2021). The electronic evolution of corporate bond dealers. *Journal of Financial Economics*, *140*(2), 368–390.

Pinter, G., and Uslu, S. (2022). Comparing search and intermediation frictions across markets. *Johns Hopkins Carey Business School Research Paper*, (pp. 22–08).

Pinter, G., Wang, C., and Zou, J. (2024). Size discount and size penalty: Trading costs in bond markets. *The Review of Financial Studies*, (p. hhae007).

Plante, S. (2021). Agency and principal trading in otc markets. *Mimeo*.

Rancière, R., and Tornell, A. (2016). Financial liberalization, debt mismatch, allocative efficiency, and growth. *American Economic Journal: Macroeconomics*, *8*(2), 1–44.

Ranciere, R., Tornell, A., and Vamvakidis, A. (2010). Currency mismatch, systemic risk and growth in emerging europe. *Economic Policy*, *25*(64), 597–658.

Rapp, A. C., and Waibel, M. (2023). Managing regulatory pressure: Bank regulation and its impact on corporate bond intermediation. *Available at SSRN 4500131*.

Rey, H. (2015). Dilemma not trilemma: the global financial cycle and monetary policy independence. Tech. rep., National Bureau of Economic Research.

Saar, G., Sun, J., Yang, R., and Zhu, H. (2023). From market making to matchmaking: Does bank regulation harm market liquidity? *The Review of Financial Studies*, *36*(2), 678–732.

Schultz, P. (2017). Inventory management by corporate bond dealers. *Available at SSRN 2966919*.

Shen, J. (2015). *Exchange or OTC market: a search-based model of market fragmentation and liquidity*. Ph.D. thesis, Chapter 1, London School of Economics and Political Science.

Stokey, N. L., Lucas, R. E., and Prescott, E. C. (1989). *Recursive methods in economic dynamics*. Harvard University Press.

Valchev, R. (2020). Bond convenience yields and exchange rate dynamics. *American Economic Journal: Macroeconomics*, *12*(2), 124–166.

Van Binsbergen, J. H., Diamond, W. F., and Grotteria, M. (2022). Risk-free interest rates. *Journal of Financial Economics*, *143*(1), 1–29.

Vayanos, D., and Weill, P.-O. (2008). A search-based theory of the on-the-run phenomenon. *The Journal of Finance*, *63*(3), 1361–1398.

Weil, P. (1989). The equity premium puzzle and the risk-free rate puzzle. *Journal of monetary economics*, *24*(3), 401–421.

Weill, P.-O. (2020). The search theory of over-the-counter markets. *Annual Review of Economics*, *12*, 747–773.

Working, H. (1949). The theory of price of storage. *The American Economic Review*, *39*(6), 1254–1262.