Three Essays in Financial Economics

Inaugural Dissertation in Fulfillment of the Requirements for the Degree of

DOCTOR RERUM OECONOMICARUM

at the

Department of Economics,

Faculty of Business, Economics and Social Sciences

submitted by

Fabienne Schneider

Advisor:	Prof. Dr. Cyril Monnet
External Examiner:	Prof. Dr. Kinda Hachem
Author's Place of Origin:	Hasle bei Burgdorf
Date of Submission:	20.06.2024

The faculty accepted this thesis on the 22 August 2024 at the request of the reviewers Prof. Dr. Cyril Monnet and Prof. Dr. Kinda Hachem as dissertation, without wishing to comment on the views expressed therein.

Die Fakultät hat diese Arbeit am 22 August 2024 auf Antrag der Gutachter Prof. Dr. Cyril Monnet and Prof. Dr. Kinda Hachem als Dissertation angenommen, ohne damit zu den darin ausgesprochenen Auffassungen Stellung nehmen zu wollen

Contents

List of Figures

List of Tables

Acknowledgements

"It takes a village to raise a child."

"It took the village of Gerzensee and its wider surroundings to raise a Ph.D."

I would like to thank

... my partner Alvaro. You have supported me every minute of my journey and even beyond. Thank you for calling me on my twentieth birthday and being with me ever since.

... my parents Jolanda & Christoph and my sister Stephanie with Yves, Valentin and my godchild Victor. You are my home and your love carries me every day.

... my supervisor Cyril. I am grateful for all the hours of help, opportunities and chances I received. You made me curious and kept me going. I am glad that it was you who accompanied me on this journey.

... my host Plamen. You encouraged me and you still do. Many thanks for your valuable advice and friendship. I enjoy working with you.

... the current directors of Gerzensee, Martin and Sylvia, as well as former director Dirk for their support in all kinds of forms throughout the entire journey and for funding my Ph.D. You provided me with an additional source of encouragement and guidance, and that was a privilege for which I am thankful.

... my co-authors Remo, Ragnar and Olav for enlightening discussions, fun and support at all stages of the projects. Many thanks, takk!

... the Norges Bank team, led by Kasper, and the BIS team, led by Leonardo, for making my external research stays possible. During these stays I met many inspiring people who shaped my view on the economics profession and the world.

... my external reviewer Kinda for reading, commenting and evaluating my thesis.

... my friend Pascale. The fun started in kindergarten and will continue for many years to come, not least thanks to my godchild Antonia.

... my office partner Lorenz. I could not imagine a better gentleman to share the same office with.

... my Gerzensee team members Teo, Manfred, Susanne, Milen, Corinne, Jacqueline, Alexandra, Nils, Daniela and Rahel. Without you, my Ph.D. years would have been much less colourful.

... my professors Fabrizio and Pierpaolo for their advice, support and motivation.

... my first year classmates Lucas and Stefanie for learning and laughing together.

... the Bank of Canada for a new adventure to come.

Introduction

This thesis consists of three separate and self-contained chapters on financial economics. The first chapter focuses on explaining anomalies in the US Treasury market and discusses a policy recommendation in the light of the current discussion on how to restructure the market after several crises. The second chapter takes a broader macroeconomic view and focuses on safe assets in general and not "just" on US government bonds. Nevertheless, similar to one of the main anomalies discussed in the first chapter (which is related to liquidity premia), we examine convenience yield differences across assets with different liquidity. We also return to the US Treasury market to test our theoretical predictions. The third and final chapter looks into the future and is more related to decentralised finance. It analyses how unsecured credit can be sustained despite anonymity (as the latter is a desire of many modern decentralised finance applications). Key is the availability of a public ledger technology, as a blockchain can provide it. In the following, I describe each chapter and its contribution and main findings in a short summary.

Chapter 1 The US Treasury market is one of the most important and liquid markets in the US financial system, is crucial for the conduct of monetary policy and is used as a benchmark around the world. Against this background, it is worrying that this market has broken down in several episodes.[1] Liquidity can evaporate quickly, as seen in the March 2020 crisis.

And even in normal times there are anomalies in the market. Every day, Treasuries with a value of USD 40 billion fail to settle.[2] In the data, I observe that this failure rate differs depending on whether a Treasury is so-called on-the-run or off-the-run. On-the-run Treasuries are all Treasuries issued at the last auction of their maturity, and off-the-run Treasuries are all others. The data shows that off-the-run Treasuries fail twice as often.[3] In addition, for the same

[1] Severe market disruptions occurred in October 2014, September 2019 and March 2020 (see U.S. Department of the Treasury et al. (2015), Anbil et al. (2020) and Schrimpf et al. (2020)).

[2] Data on settlement fails are provided by the Depository Trust and Clearing Corporation (DTCC) and can be downloaded here: https://www.dtcc.com/charts/daily-total-us-treasury-trade-fails.

[3] Separate data on fails in on- and off-the-run Treasuries are reported in the NY FED's Primary Dealer Statistics.

cash flow and maturity, they have to pay a much higher yield and are traded in much lower volumes than on-the-run Treasuries.[4] This yield anomaly is called the on-the-run premium (see Vayanos and Weill (2008) and D'Amico and Pancost (2022)).

In the first part of the first chapter, I document these anomalies and explain them using a model in which dealer inventory risk is key. My work is complementary to others on these anomalies such as Vayanos and Weill (2008). Compared to the model of Vayanos and Weill (2008), which has a premium on either the on-the-run or the off-the-run asset (depending on the equilibrium), in my model the premium is always on the on-the-run asset.[5] In a second step, I use my model to analyse one of the recommendations on how to restructure the market. The debate on whether the market structure needs to be redesigned was catalysed by the recent market crisis in March 2020. Off-the-run Treasuries were at the epicentre of the crisis (Wells (2023)). It is therefore interesting to distinguish the impact on on- and off-the-run assets separately. I analyse the impact of broad access to central bank facilities. Due to the growing size of the Treasury market and the regulatory costs of large bank balance sheets, more and more non-banks are active in the Treasury market. This calls into question any superior access of primary dealers to central bank facilities.

A key role in my model is played by sellers active in the Treasury market, who short certain Treasuries (i.e. on- or off-the-run Treasuries) to buyers of them. When a Treasury is on-the-run, all primary dealers have similar stocks of it (as it has just been issued) and shortsellers can locate them. On the other hand, the inventories of off-the-run Treasuries held by primary dealers vary according to their past trading history. The trading history is influenced by the over-the-counter frictions of the Treasury market. Varying inventories imply uncertainty about the amount of assets available when short sellers trade with primary dealers. This leads to a higher frequency of settlement fails for off-the-run assets. This makes them unattractive, leading to lower trading volumes, and they have to pay a higher yield, resulting in the on-the-run premium.

I also test my theory empirically using data from the NY Fed's Primary Dealer Statistics. Consistent with my theory, I find a positive correlation between on-the-run premia and settlement fails in off-the-run assets and a negative correlation between the latter and primary dealer inventories.

In the second part of this chapter, I show that broad access to a reverse repo facility stimulates trading. The on-the-run premium decreases and I present empirical evidence consistent with this finding. On the other hand, settlement fails increase and, surprisingly, it's the primary

[4]The TRACE data reports trading volumes in on-the-run and off-the-run Treasuries. The TRACE data is available here: https://www.finra.org/finra-data/browse-catalog/about-treasury/monthly-file.

[5]The reason is that I explicitly take into account the different time the assets have been in the market since issuance.

dealers who benefit from wider access, not those who gain access.

Chapter 2 The US government bonds I discussed in the previous chapter are so-called safe assets. But as we have just seen, their valuation depends on more than just their safety. There are other reasons for holding safe assets, such as the fact that they are often highly liquid. A large literature discusses safe assets and convenience yields, centred around a seminal contribution by Krishnamurthy and Vissing-Jorgensen (2012).[6] Differences in convenience yields can be substantial even when the assets are very similar. The second chapter of this thesis, co-authored with Ragnar Juelsrud, Plamen Nenov and Olav Syrstad, focuses on this aspect. We analyse theoretically and empirically the cross-section of convenience yields on safe assets. We show how it adjusts and affects welfare in response to changes in various factors such as overall or compositional changes in supply and aggregate market risk (e.g. during flight-to-safety periods). Our results can also be used to assess the impact of central bank operations such as quantitative easing and operation twists on welfare. Importantly, in all our results we discuss the direct effects on the targeted assets but, more interestingly, also the spillovers to all other assets.

We use a tractable asset pricing model with multiple safe assets and a risky benchmark portfolio. Agents face liquidity risk and can only self-insure by holding and trading assets. The assets differ in their transaction costs. The convenience yield is given by the return of any safe asset with positive transaction costs relative to the return of the most liquid (zero transaction cost) safe asset in the economy. As a first step, we decompose the convenience premia into their underlying drivers. We identify an aggregate price of convenience in the economy. Any change in this will affect the convenience yields of all safe assets. However, the price of convenience is not "only" a pricing factor. It also measures the equilibrium liquidity scarcity in this economy relative to a first-best full liquidity insurance. The price of convenience is therefore a sufficient statistic for this economy's deviations from first-best.

We use our framework to discuss the impact of changes to the "pool of safe assets". We look at changes in the composition and overall size of the pool. For example, changing the composition of the pool by substituting less liquid safe assets for more liquid ones reduces the overall price of convenience, shrinks the convenience yields on all safe assets and increases welfare. The opposite substitution works in the opposite direction. We refer to this effect as

[6]Related papers include Krishnamurthy and Li (2023), Nagel (2016), Caballero et al. (2016) and Krishnamurthy and Vissing-Jorgensen (2015), to name a few. The literature on transaction costs and trade frictions is also related, as will become clearer later. See for example Vayanos and Vila (1999),Vayanos (2004) and Vayanos and Vila (2021). Finally, we use differences in the Treasury-OIS spreads across assets as a measure of convenience yields. Other papers using the Treasury-OIS spread are He et al. (2022), Du et al. (2023) and Klingler and Sundaresan (2023).

the “purifying effect” (and “polluting effect” for the opposite substitution). Therefore, the type of safe asset that is substituted or added to an economy matters for aggregate liquidity and welfare. As central bank interventions, such as operation twist and quantitative easing, also affect the economy by changing the composition and overall supply of safe assets, our analysis can be used to assess the asset pricing and welfare effects of these interventions.

Finally, we find a safety value channel during periods of increased aggregate risk or risk aversion. This is interesting in terms of flight-to-safety periods. Due to a revaluation effect of risk-free assets, all convenience premia decline when such risks increase.

In the second part of the chapter, we then test our theoretical predictions using data from the US Treasury market. We use the difference in the Treasury-OIS spreads[7] between longer term Treasuries and 3-months Treasury bills (a double difference) as the counterpart to the convenience yield between safe assets in our model. To test the predictions for the supply of safe assets, we use two sources of variation: First, we use changes in primary dealer positions as a proxy for the supply of Treasuries of different maturities to market participants. Second, we use debt ceiling episodes to obtain exogenous variation in the supply of Treasuries. Finally, for our other results, we use the VIX as a measure of risk and the MOVE Index as a proxy for transaction costs, as it measures the yield volatility of US Treasuries and is thus closely correlated with illiquidity (Duffie et al. (2023)). Our theoretical results are confirmed by the data.

Chapter 3 There is a widespread belief that credit cannot be anonymous in the absence of collateral. It is assumed that knowledge of the borrower’s identity is necessary to punish him in case of default. In the third chapter of this thesis, co-authored with Remo Taudien, we show the existence of an anonymous credit equilibrium. Anonymity comes in different forms. We apply the concept of pseudonymity, where agents use, for example, wallet addresses to trade while keeping their identities hidden. Pseudonymity lies between the two extremes of strict anonymity and full information. This form of anonymity is interesting because many emerging decentralised finance (DeFi) applications are built on it, emphasising the importance of anonymity. They often use blockchain technology, as it allows for the maintenance of a public ledger of all past activities of wallet addresses, which is necessary to maintain pseudonymity. With this in mind, it is surprising that there are not many studies analysing pseudonymous credit. Relevant papers on credit, pseudonymity or endogenous credit limits are for example Friedman and Resnick (2004), Wang and Li (2023) and Alvarez and Jermann (2000).[8]

[7] OIS stands for overnight indexed swap.

[8] As we will combine credit with reputation accumulation (see the discussion below), the “starting small”

In our model, there are borrowers and lenders. Borrowers use accounts to trade with lenders, and only their actions (as borrowing and repaying) are revealed to each other and to the public (not the identities of the borrowers). Borrowers and lenders meet bilaterally "through the accounts" in a first of two subperiods. The lender produces a good that the borrower is unable to produce and wants to consume. The borrower cannot produce anything for the lender in this subperiod and can therefore only promise to repay in the next subperiod. Trade in this economy is therefore based on credit. A key feature of our model is that a previous default can always be "hidden" by opening a new account, which is possible at zero cost. Moreover, the opening of a new account does not necessarily have to be due to a previous default, since new entrants to the economy also open up new accounts.[9] In our main result, we show that credit can still be sustained. We show how to build up reputation schemes that allow for credit and always exist. Key is that accounts with higher reputations are rewarded with more credit. This makes default costly, even though new accounts (i.e. pseudonyms) can be created at zero cost. However, such a pseudonymous credit system is not without cost. Since each new account holder starts with a small amount of credit due to low reputation, new entrants will initially also receive a small amount of credit and therefore have a low consumption rate. An interesting side result of our work is that although agents are allowed to have two accounts with positive reputation in parallel (and as many as they want in sequence), they only use one. Intuitively, using a second account would mean forfeiting the chance to consume with the account that already has a higher reputation. Finally, we show that instead of free accounts and building reputation over time, making accounts costly can also prevent default and maintain a pseudonymous credit equilibrium. This last example has a high technological feasibility and applicability.

Our thought experiment is a first step towards a future with a decentralised financial system, where agents can extend unsecured credit to each other while maintaining their anonymity.

literature (see Hua and Watson (2022) for an example) is also related. From a model perspective, the new monetarist literature is relevant (see Lagos et al. (2017) for a review of this literature).

[9] In our model, borrowers leave the economy with an exogenous probability and are replaced by new borrowers.

First Chapter

On-the-run Premia, Settlement Fails, and Central Bank Access

On-the-run Premia, Settlement Fails, and Central Bank Access[1]

Abstract

The premium on "on-the-run" Treasuries (i.e. the most recently issued ones) is an anomaly. I explain it using a model in which primary dealers hold inventories of Treasuries. Primary dealers are more likely to hold large inventories of on-the-run Treasuries. There is also less variation across primary dealers in the available stock of on-the-run Treasuries compared with all other, so-called off-the-run Treasuries. Because on-the-run Treasuries are easier to find, they trade at a premium. My theory is consistent with the USD 40 billion of Treasury contracts that fail to settle each day, with the median failure rate of off-the-run Treasuries being almost twice that of on-the-run Treasuries. I use the model to analyse the effects of granting access to central bank facilities to non-banks active in the Treasury market. Broad access stimulates trading and reduces the on-the-run premium, but settlement fails increase and, counterintuitively, only primary dealers benefit.

[1] I thank Bruno Biais, Martin Brown, Jens Christensen, Stefania D'Amico, Michael Fleming, Refet Gürkaynak, Arvind Krishnamurthy, Ricardo Lagos, Antoine Martin, Cyril Monnet, Plamen Nenov, Dirk Niepelt, and the participants of the Macroeconomics PhD Seminar of the University of Bern (2021), Central Bankers Course on Money Markets, Liquidity, and Payment Systems (2022), Macroeconomics Workshop Hasliberg (2022), BDP Alumni Conference (2022), Brown Bag Seminar of the University of Bern (2022), 35th Australasian Finance & Banking Conference (2022), Swiss Society for Financial Market Research Conference (2023), the seminar at the Bank of Canada (2024), the Bank of England (2024), and the Sveriges Riksbank (2024) and other seminars and conferences for feedback, insights, and discussion.

3.1 Introduction

With an average daily trading volume of half a trillion US dollars, the US Treasury market is one of the most important and liquid markets in the US financial system, crucial to the conduct of monetary policy and a key pillar of the US economy. Despite its importance, the US Treasury market exhibits some irregularities, which I describe in detail in the next section and summarise here.[1]

First, it is well known that on-the-run Treasuries – the most recently issued Treasuries – trade at significantly lower yields, higher prices and lower repo rates than other Treasuries (known as off-the-run) with similar cash flows and maturity dates, giving rise to a puzzling arbitrage opportunity known as the "on-the-run" premium (see Vayanos and Weill (2008), D'Amico and Pancost (2022) and figure 3.2a in the next section).[2] Second, despite trading at a premium, the volume of trades in on-the-run Treasuries is much larger than that in off-the-run Treasuries (see figure 3.2b in the next section). Third, on average USD 40 billion of Treasury contracts fail to settle each day (see figure 3.3a in the next section). Fourth, interestingly, failure rates differ by Treasury type, with on-the-run Treasuries having a median settlement failure rate almost half that of off-the-run Treasuries (see figure 3.3b in the next section).

These stylised facts and irregularities raise the following questions: How can there be a premium on certain Treasuries and why is it always on the on-the-run Treasuries? Why do the cheaper off-the-run Treasuries trade at lower volumes? How can there be settlement fails in a benchmark market such as the US Treasury market, and why do the off-the-run Treasuries fail to settle more often?

In the first part of the paper, I develop a model of the US Treasury market to answer these questions. In the second part, which I describe in more detail below, I use it to conduct policy analysis motivated by current discussions about how to restructure the market.

The US Treasury market model incorporates the key features of the market that I describe in section 3.2: It is an over-the-counter (OTC) market where primary dealers are the first acquirers of Treasuries at the primary auction. I assume that there are three types of agents: sellers, buyers and primary dealers. A seller is any financial entity other than a primary dealer, such as a non-bank, that sells Treasuries short. A buyer is akin to a long-term holder of Treasuries, such as a pension fund. There are different types of Treasuries, on- and off-the-run, and to simplify the model I do not model the primary auction of Treasuries, but assume that primary

[1]The average daily trading volume is the volume reported to TRACE between February and October 2023. The TRACE data is available here: https://www.finra.org/finra-data/browse-catalog/about-treasury/monthly-file.

[2]See figure 3.12.4 in the appendix for a graphical representation of the on-the-run cycle.

dealers are endowed with the latest issue of Treasuries.[3] Buyers have the highest valuation for Treasuries, but the market is segmented and they cannot contact primary dealers directly, only through sellers.

In a first market, sellers sell financial contracts to buyers that promise to deliver a specific type of Treasury (for short, on- or off-the-run, including maturity date). Because the seller can fail to settle, the contract is secured by collateral. Next, sellers contact primary dealers to buy the desired type of Treasury in an OTC market. There, the seller is randomly matched with a primary dealer. The primary dealers always have the most recent issue of Treasuries in their inventory, as they have just been auctioned. However, depending on their trading history, they may not have enough of the desired off-the-run Treasuries. In this case, the seller fails to settle and he delivers as many Treasuries as possible to the buyer in accordance with their contracts. If necessary, the buyer seizes the collateral to cover the undelivered amount. These fails do not occur with on-the-run Treasuries because all primary dealers hold the same inventories since they were just filled up. Once all the trades have been conducted, the sellers can deposit any remaining idle balances in a central bank facility and receive an interest rate on them.[4] Sellers then go into the next sequence of trades.

In equilibrium, I show that the occurrence of settlement fails leads to a preference for the safer on-the-run Treasuries. Because they have been in the market for a shorter time, sellers are more likely to find them and they have a greater chance to settle. Therefore, on-the-run Treasuries trade at a premium and in greater volume than off-the-run Treasuries with the same cash flow and maturity date, explaining the stylised facts for the US Treasury market mentioned in the beginning. I also provide empirical evidence that lower inventories imply more settlement fails in off-the-run Treasuries, and more such settlement fails imply higher on-the-run premia, as predicted by the model.

In a second part of the paper, I use the model to shed light on the current policy discussion about the need to restructure the Treasury market (see the discussion at the Jackson Hole conference by Duffie (2023)). The background to this discussion is epitomised by the US Treasury market crisis of March 2020. In the aftermath of the great financial crisis of 2007, primary dealers faced tighter regulatory constraints, leading them to reduce their balance sheet space for Treasuries. At the same time, the US Treasury market grew strongly. Non-bank financial

[3]In the extension in section 3.10 the auction is included.

[4]The facility in my model can be interpreted as a deposit facility or a reverse repo facility where I focus on the cash leg and abstract from the collateral part. First, the facility's repos are general collateral repos and the cash lender is willing to receive any security that falls into a broad class. He does not search for a specific security (Bowman et al. (2017)). Second, even if the facility were to provide a specific security that was sought, the security would have to be returned the next day and the facility would only provide temporary availability.

institutions have filled the space left by primary dealers.[5] But in March 2020, the presence of non-banks in the US Treasury market led to a rapid drying up of liquidity and a sharp decline in market depth, exacerbated by the reluctance of primary dealers to take more US Treasuries onto their balance sheets (Eren and Wooldridge (2021)). Off-the-run Treasuries were at the epicentre of the crisis (Wells (2023)). This is reflected in the on-the-run premia across all maturities, which rose sharply as shown in figure 3.1.[6]

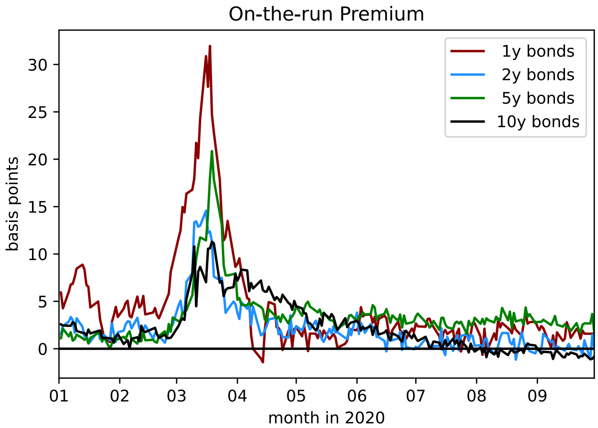


Figure 3.1: On-the-run premia during the "March 2020" breakdown

Among their ten recommendations for the US Treasury market, the Working Group on Treasury Market Liquidity led by Duffie, Geithner, Parkinson and Stein recommends broad access to central bank repo financing (Duffie et al. (2021)). They criticise the current facility for providing limited access to primary dealers and banks rather than a broad range of market participants.[7]

The model helps to understand the impact of broad access to central bank facilities on prices, premia, traded quantities, fails and profits (in a general setting in normal times). A first observation is that the facility is not a substitute for trading, but rather complements it. In this model, the facility (like a deposit or reverse repo facility) generates a certain return on liquid funds between trades for those who have access. This feeds back into the overall cost of trading. If sellers gain access, an increase in the facility rate stimulates trading and prices rise. The stimulated trading implies that more Treasuries end up in the hands of buyers and less in

[5] On the BrokerTec platform, one of the main marketplaces, non-banks, especially principal trading firms, already accounted for more than half of the trading in benchmark 5-year, 10-year and 30-year bonds in 2015. Traditional banks and dealers had a share of 30-40%.

[6] As in Christensen et al. (2017), the on-the-run yield is subtracted from the par yield of seasoned bonds. The data are taken from the FED yield curve, which is an updated version of the original Gürkaynak-Sack-Wright curve (Gürkaynak et al. (2010)), https://www.federalreserve.gov/data/tips-yield-curve-and-inflation-compensation.htm, and the FRB-H15 tables, https://www.federalreserve.gov/releases/h15/.

[7] Another example is the FED's reverse repo facility. An increasing number of institutions already access this facility (Frost et al. (2015), Baklanova et al. (2015) and Marte (2021)).

the inventories of primary dealers. Settlement of off-the-run Treasuries is more likely to fail. The reason is that as more Treasuries are sold early, fewer are available during the off-the-run period.

Interestingly, as I found in the data, the premium decreases as the facility rate rises. The reason is intuitive: as off-the-run Treasury prices initially rise, so does the value of the collateral. This leaves buyers better off in the event of default, and buyers increase their demand for the contract with the off-the-run Treasuries. Therefore, the increase in the price of the off-the-run Treasuries is greater than the increase in the price of the on-the-run Treasuries. Also because of this additional demand effect, the increase in the quantity of off-the-run Treasuries traded is initially larger than the increase in the quantity of on-the-run Treasuries.[8]

Paradoxically, in equilibrium, only the primary dealers benefit from a rise in the facility rate, and those who are granted access do not. This is because the primary dealers can now sell more Treasuries at a higher price. By contrast, perfect competition in the contract market erodes any advantage that sellers may have. Finally, buyers of Treasuries lose, first through higher prices and second through an externality. The buyer does not take into account that if he buys more Treasuries early, fewer will be available during their off-the-run period, implying a higher default rate.[9]

Related Literature First and foremost, my paper is related to the literature on the on-the-run premium. A well-known framework for the on-the-run premium is provided by Vayanos and Weill (2008). They have a setup where there are two assets with identical cash flows and agents can go long or short an asset. Since short sellers have to deliver the asset they have borrowed, they face search externalities and favour the asset that is more liquid. Liquidity is self-fulfilling in their model. As in Vayanos and Weill (2008) I also include OTC market frictions and delivery constraints in my model. However, my model is dynamic whereas theirs is static. In addition, I include a key factor to explain the on-the-run premium: the fact that one asset, the off-the-run asset, has been available in the market for a longer time. This also allows for equilibrium selection with a premium on the on-the-run asset, which is not the case in Vayanos and Weill (2008), who have two self-fulfilling equilibria with the premium on either asset.

Another theory of the on-the-run premium comes from Pasquariello and Vega (2009). In their model, the premium arises from endowment shocks. There are two frictions: information heterogeneity and imperfect competition among traders. My model is similar to theirs in the sense that I include uncertainty due to limited information. In my model, the uncertainty

[8]This policy effect is particularly relevant in the context of the recent crisis, where the market for off-the-run Treasuries froze (Eren and Wooldridge (2021)).

[9]Few of the results depend on search frictions being above a small minimum. See appendix 3.12.2.5.

is about the stock of off-the-run Treasuries. In their model, it is with respect to on-the-run Treasuries, since the endowment shocks received by agents are private information.

Broadly speaking, compared to Vayanos and Weill (2008) and Pasquariello and Vega (2009), my model focuses on primary dealer inventory and intermediary settlement risk to explain the premium. All dynamics are driven solely by the fact that Treasuries are in the market for a different length of time since issuance. This element implies that the premium is always on the on-the-run asset. The model allows for a more general discussion of the US Treasury market and, in addition, I can also analyse the impact of central bank facility access.

Empirical work attempting to explain the on-the-run premium includes, for example, Strebulaev (2002), Goldreich et al. (2005) and D'Amico and Pancost (2022). Strebulaev (2002) suggests that the premium may measure differences in tax treatment rather than liquidity premia. Goldreich et al. (2005) distinguish between current and future liquidity and suggest that expected future liquidity, not just current liquidity, determines prices and is a significant driver of the on-the-run premium. D'Amico and Pancost (2022) link the on-the-run premium to the risk of unexpected fluctuations in the collateral value of Treasuries.

A recent paper related to mine is Corradin and Maddaloni (2020). They build on Vayanos and Weill (2008) and study central bank intervention. Compared to my paper, they do not study access to facilities but central bank purchases. Specifically, they analyse how purchases by the European Central Bank affected repo specialness in the Italian government bond repo market during the euro area sovereign debt crisis. Specialness is the premium paid to procure a particular security in the repo market. On-the-run securities often trade as "special". Important early work on this topic was done by Duffie (1996) and Krishnamurthy (2002). Corradin and Maddaloni (2020) show that purchases reduce liquidity and increase specialness in the presence of short selling. They also show that assets in high demand and older assets with lower turnover are more likely to fail. The probability of default increases with the specialness of the asset. In contrast, Liu and Wu (2017) show that the on-the-run premium is low when counterparty risk is high. Compared to my model, the risk in Liu and Wu (2017) and Corradin and Maddaloni (2020) refers to variations in a general risk measure or only in the specific asset. Corradin and Maddaloni (2020) focus exclusively on crisis periods, Liu and Wu (2017) state that their results are particularly pronounced in such periods.

Second, my paper adds to the literature on settlement failures (see, e.g., Fleming et al. (2014), Fleming and Garbade (2002), Fleming and Garbade (2004), Fleming and Garbade (2005), and Garbade et al. (2010)) as well as on non-banks in the Treasury market and their access to the reverse repo facility (see, e. g. Eren and Wooldridge (2021), Doerr et al. (2023), and Frost et al.

(2015)). More generally, my work examines the microstructure of the Treasury market and speaks to the literature discussing how to reform it (see, e.g. Duffie (2020), Duffie et al. (2021), Duffie (2023), Durham and Perli (2022), Fleming and Keane (2021), He et al. (2022), Schrimpf et al. (2020), and Vissing-Jorgensen (2021)). The literature on the OTC market environment and dealer markets is also related (see e.g. Duffie et al. (2005), Huh and Infante (2021), Lagos and Rocheteau (2009), and Li and Schürhoff (2018)).

The structure of the paper is as follows: Section 3.2 describes the US Treasury market, section 3.3 the environment and section 3.4 the value functions. The equilibrium is also defined. Section 3.5 proves the existence of our main equilibrium of interest. Section 3.6 discusses how the model explains the stylised facts and why the premium is always on the on-the-run Treasury. Section 3.7 empirically analyses the relationship between the on-the-premium, the settlement fails, and the primary dealer inventories in the data and tests theoretical predictions. The implications of broad access to central bank facilities are analysed in section 3.8. Given the theoretical results, section 3.9 analyses the dependence of the on-the-run premium on the reverse repo facility rate in the data. Section 3.10 presents the Treasury life cycle. Section 3.11 concludes.

3.2 Description of the Treasury market

In this section I describe the US Treasury market, its structure and trading dynamics, and provide evidence for the stylised facts highlighted in the introduction.

The Treasury spot market is OTC (Fleming et al. (2018)). This means that there is no all-to-all trading at a central venue and no central pricing. Depending on the security traded and the trading partners involved, the degree of friction in the OTC market varies. For example, dealer-to-dealer trading of benchmark on-the-run Treasuries on electronic platforms such as BrokerTec is less frictional than interdealer and dealer-to-customer trading of the less liquid off-the-run Treasuries intermediated on voice and more manually assisted electronic platforms (Bessembinder et al. (2020) and U.S. Department of the Treasury et al. (2015)).[10]

On-the-run Treasuries are the most recently issued Treasuries of a given maturity, and all previously issued Treasuries of the same maturity are referred to as off-the-run. As figure 3.2a shows for 10-year Treasuries, on-the-run Treasuries trade at significantly lower yields than off-the-run Treasuries with very similar cash flows and maturity dates.[11] They also have higher prices and lower repo rates. The term "very similar" refers to the fact that, in general, there

[10]The overall share of trading on all types of electronic platforms in the US Treasury market is 70 percent (Bech et al. (2016)).

[11]The data are taken from the FED yield curve, which is an updated version of the original Gürkaynak-Sack-Wright curve (Gürkaynak et al. (2010)), https://www.federalreserve.gov/data/tips-yield-curve-and-inflation-compensation.htm, and the FRB-H15 tables, https://www.federalreserve.gov/releases/h15/.

are not two Treasuries in the market with exactly the same cash flow and maturity date where one is on-the-run and the other is off-the-run. In fact, if you want to compare Treasuries with the same overall maturity, it is impossible to do so. In practice, therefore, one either compares Treasuries with the same overall maturity using an estimated off-the-run yield curve (see, for example, the first figure in Christensen et al. (2017)), or one abstracts from small differences in cash flows and maturity dates, or one compares on- and off-the-run Treasuries with the same maturity date that have a different overall maturity (see, for example, Christensen et al. (2020)). In figure 3.2a, as in Christensen et al. (2017), the on-the-run yield is subtracted from the par yield of seasoned bonds.

Also, despite being more expensive and far fewer in number, on-the-run Treasuries trade in much larger volumes than off-the-run Treasuries, as shown in figure 3.2b.[12] This is true whether I look at dealer-to-customer or interdealer and automated trading system (ATS) trades.

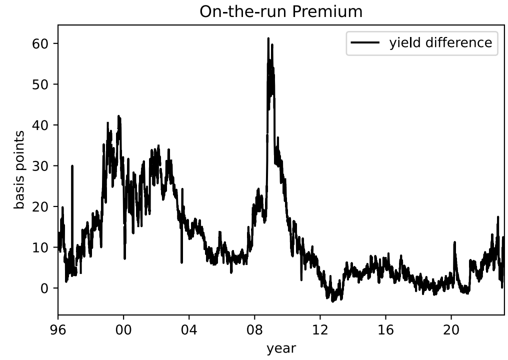
Another surprising fact is that, on average, USD 40 billion of Treasuries are not delivered on time to settle a contract each day.[13] These events are commonly referred to as "settlement fails". Figure 3.3a shows the failure rate, which is calculated by dividing the value of Treasuries that failed to be delivered on time by the value of all Treasuries traded. Interestingly, the failure rates differ depending on whether a Treasury is on- or off-the-run, and figure 3.3b shows that fails involving on-the-run Treasuries are less frequent than those involving off-the-run Treasuries.[14]

The OTC structure implies that there are search costs that can explain settlement fails. In particular, search costs become relevant when financial contracts include delivery constraints. For example, spot market trades are often complemented by special repo trades to short-sell specific Treasuries. "Special" refers to the fact that the collateral of the repo is fixed and determined by its number, called ISIN or CUSIP, and the repo may have a rate that differs from the general collateral rate. To short-sell a particular Treasury, it is borrowed using a special repo and sold in the market today. The next day, a Treasury with the same ISIN or CUSIP is bought in the spot market, preferably at a lower price than it was sold on the previous

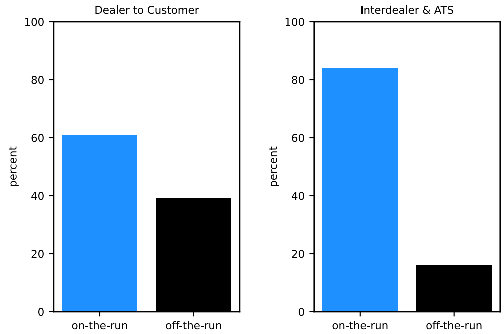
[12]Trading volumes in on-the-run and off-the-run Treasuries are the volumes reported to TRACE between February and October 2023. The TRACE data is available here:
https://www.finra.org/finra-data/browse-catalog/about-treasury/monthly-file.

[13]The data is provided by the Depository Trust and Clearing Corporation (DTCC) and can be downloaded here: https://www.dtcc.com/charts/daily-total-us-treasury-trade-fails. In times of stress, the daily value can spike. Current policy discussions consider central clearing as an effective way to significantly reduce fails in the future (Fleming and Keane (2021)). For more information on settlement fails, see Fleming and Garbade (2005).

[14]The data are from the FED's primary dealer statistics. It includes outright and financing fails. The median failure rate for on-the-run Treasuries is 0.36% and the median failure rate for off-the-run Treasuries is 0.66%. The rates are not an exact measure. This is because one part of the time series used in the calculation is an average over the reporting week and the other part of the time series reports a value as of the reporting weekday. Given the high frequency, this should not matter and the observed pattern is clear. Each series is outlier adjusted, where an outlier is defined as being below the 2.5% percentile and above the 97.5% percentile. Rates up to 2.5% are shown in the figure. Few rates are higher.

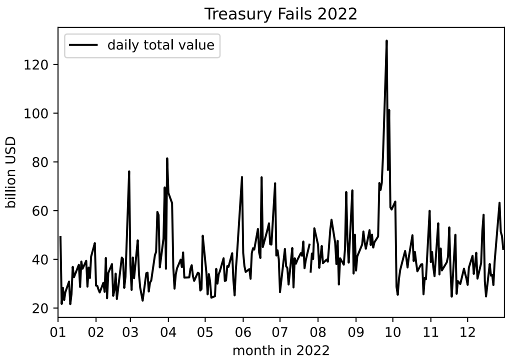


(a) On-the-run premium on 10 year Treasury bonds

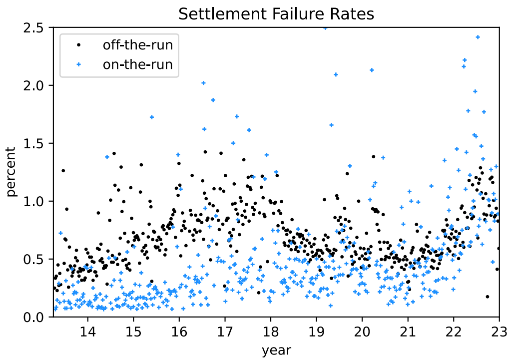


(b) Trading volumes

Figure 3.2: Yields and volumes



(a) Daily settlement fails



(b) Fails of on- and off-the-run Treasuries

Figure 3.3: Fails

day, and returned to the lender in the repo transaction. If such a Treasury cannot be found, a settlement fail occurs. The borrower of the Treasury in the repo transaction pays a penalty, the Treasury Market Practice Group (TMPG) fail charge.[15] Fleming and Keane (2021) write that on-the-run Treasury fails account for less than a quarter of all fails in non-crisis periods.[16] The figure in Fleming et al. (2014), presented in the appendix 3.12.3, shows that gross fails are much higher in seasoned Treasuries than in others (including on-the-run Treasuries).

Note that arbitrage to exploit the price difference between on- and off-the-run Treasuries involves short selling, but is mostly prohibited by efficient markets because repo rates for them also differ (Krishnamurthy (2002)).

[15] For more information on the TMPG fail charge, see https://www.newyorkfed.org/tmpg and Garbade et al. (2010).

[16] In addition, they also note that on-the-run Treasuries are more often involved in so-called daisy chain fails. One fail implies another as the trades are linked in a chain.

3.3 The environment

Time is discrete and goes on forever, $t = 0, 1, ..., \infty$. The discount factor is $\beta \in (0, 1)$ and each period consists of two subperiods. There are three types of infinitely lived agents in the model: a buyer, a seller, and a primary dealer.[17] There is a continuum in each type.

There are two segmented sequential markets. The first market is called the spot market. It takes place in the first subperiod and is an OTC market. The second market is Walrasian and takes place in the second subperiod. It is called contract market. Figure 3.4 gives an overview over the timeline.

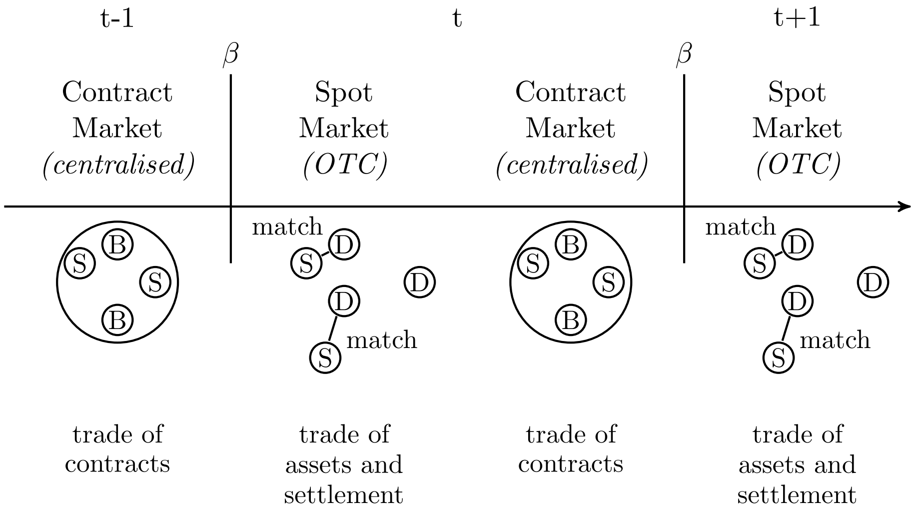


Figure 3.4: Timeline

There are two goods: a settlement good and real coupons. The settlement good $m \in \mathbb{R}_0^+$ is storable and divisible. Coupons δ are perishable and are given by two assets: one asset gives one coupon per each second subperiod for two consecutive periods, the other for one period. Assets are storable and divisible. An asset is on-the-run if it belongs to the most recently issued generation of its maturity. Therefore, in each period there are two types of on-the-run and one type of off-the-run assets available for trading: the two-period assets maturing in two periods (2), the two-period assets maturing in one period (f), and the one-period assets maturing in one period (n). The letters n and f refer to their respective on-the-run (n) and off-the-run (f) state. I will refer to them henceforth as the on-the-run asset and the off-the-run asset. The two-period asset maturing in two periods is also on-the-run, but its state is not relevant to the analysis.[18] Figure 3.5 gives an overview over the assets and their cash flow.

[17]A seller is any financial entity other than a primary dealer. It can for example be a non-bank. A buyer can be interpreted as a long-term holder (e.g. a pension fund).

[18]Only in very few contexts both types of on-the-run assets are meant when using the term on-the-run.

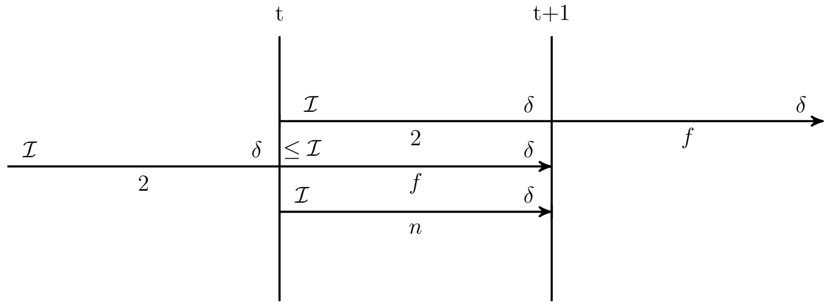


Figure 3.5: Assets in period t

The figure illustrates that the on-the-run (n) and off-the-run (f) assets are identical in terms of maturity date and coupon. The only difference is the issuance date. Therefore, I compare these two assets to measure the on-the-run premium.[19] At the beginning of each period, primary dealers are endowed with a stock $I^2 = \mathcal{I} \in \mathbb{R}^+$ of newly issued two-period assets and a stock $I^n = \mathcal{I} \in \mathbb{R}^+$ of newly issued one-period assets. The primary dealers' stock I^f of off-the-run assets is endogenous. I assume that buyers and sellers have knowledge of the primary dealers' inventory stock of newly issued assets. They also know the distribution of primary dealers inventories of off-the-run assets (as they know the matching probability described below), but not the inventories of each primary dealer.

All agents have linear utility δ from consuming the coupons in the second subperiod. Buyers additionally receive utility g each period when holding an asset.[20] Sellers do not value the coupons.[21] All agents have linear utility (disutility) from consuming (producing) the settlement good. The seller can produce it only in the second subperiod. The others can always produce it.[22] It is used to settle trades. It has properties similar to money except that it is a real asset.[23]

In the OTC spot market, primary dealers and sellers trade assets. Primary dealers have a match with probability $(1-\sigma) > 0$ with a seller. Only matched primary dealers can trade. Sellers always have a match with a primary dealer.[24] Buyers have no access to this market.

[19]The two assets have the same cash flow to maturity and the same maturity date as in Vayanos and Weill (2008) and Pasquariello and Vega (2009). See also, for example, Christensen et al. (2020) for how the premium can be measured. Note that if I would use the two-period on-the-run asset to calculate the premium, I would have to abstract from the second coupon rate as cash flow differences should not be the reason for the premium. The premium and all other results would remain the same.

[20]g can be interpreted as a hedging benefit from holding the asset or simply as a different valuation.

[21]It would not change any results if sellers would value the coupons as much as primary dealers.

[22]This aspect of the model ensures that there is no incentive for the seller to build up settlement goods only to deposit them in the facility. An agreement between a seller and a buyer or a primary dealer whereby the buyer or primary dealer would produce settlement goods for the seller so that the seller could deposit it and pay it back later with a profit is not possible because the seller cannot commit.

[23]The only difference from a nominal model is that the settlement good does not lose or gain value over time due to inflation. The dynamics would not change with a nominal model, and therefore discussing inflation would not add anything relevant.

[24]I make this assumption for simplicity. Changing it would not change the dynamics.

Primary dealers sell quantities of assets A^i to sellers at price p^i, where $i = \{2, f, n\}$. They face an adjustment cost of $\kappa(A^i)$ when selling assets i in terms of settlement good.[25] The function introduces a non-linearity into the model and leads to an interior solution and a determined price.[26] I assume that $\kappa(0) = 0$, $\kappa'(A^i) > 0$, $\kappa''(A^i) > 0$, and that the function is continuous.[27] Possible interpretations of the function are a nonlinear portfolio adjustment cost (see Gârleanu and Pedersen (2013) and Bacchetta and van Wincoop (2021) for examples) or a regulatory cost (see, for example, Macchiavelli and Pettit (2021)).

In every other subperiod so-called contracts are traded by the seller and the buyer. The market is called contract market. The seller sells the contracts to the buyer. A contract is a list $l^i = [a^i, \omega^i, q^i]$. a^i specifies the amount of assets i promised to be delivered in the next subperiod. Sellers cannot commit. $\omega^i a^i$ is the collateral (in settlement good) that the seller has to post at the moment of selling the contract. The buyer has first claim over the collateral in the event of non-delivery.[28] q^i is the contract price in the second subperiod. $q^i a^i$ is the payment (in settlement good) due from the buyer at settlement in the next subperiod. Figure 3.4 gives an overview over the markets.

In the basic model the seller has access to a central bank facility. The facility can be accessed every first subperiod for one subperiod. Sellers can deposit settlement good and receive an interest rate r_t on it. I assume that $\beta(1 + r_t) < 1$. This implies that it is not worthwhile to accumulate settlement good one period in advance in order to deposit it in the facility.[29]

3.4 Value functions and equilibrium

3.4.1 Primary dealer

The primary dealer value function at the beginning of the contract market, when holding asset inventories I_t^2, I_t^f, and I_t^n is

$$V^D(I_t^2 = \mathcal{I}, I_t^f, I_t^n = \mathcal{I}) = \beta(1-\sigma)\Big\{\sum_i \Big[p_t^i A_t^i - \kappa(A_t^i) + \delta(I_t^i - A_t^i)\Big] + \beta V^D(\mathcal{I}, \mathcal{I} - A_t^2, \mathcal{I})\Big\}$$
$$+ \beta\sigma\Big\{\sum_i \delta I_t^i + \beta V^D(\mathcal{I}, \mathcal{I}, \mathcal{I})\Big\}.$$

[25] In equilibrium they only sell assets.

[26] The main results also hold without this function. It gets relevant when I discuss access.

[27] Another possibility would be that the function depends on the sum of all assets sold. But I can show that for a positive premium this cannot be the case. Also, I need that the function is increasing and convex.

[28] The collateral in the form of settlement good is similar to the Treasury Market Practice Group (TMPG) fails charge. This fee allows a buyer of Treasuries to claim monetary compensation from the seller if the seller fails to deliver the Treasuries on time. For more information, see https://www.newyorkfed.org/tmpg and Garbade et al. (2010).

[29] I could in addition assume that primary dealers also have access to the facility, but this does not change the dynamics. This is why I am omitting it to ease notation.

With a probability of $(1-\sigma)$, the primary dealer has a match with a seller in the spot market and can sell assets. Optimal prices and quantities are determined by the bargaining problem (3.2) described below. For each kind of asset $i \in \{2, f, n\}$, the primary dealer sells the optimal amount A_t^i. When selling the amount A_t^i he receives the price p_t^i for each of them and faces the cost $\kappa(A_t^i)$. In addition, he cannot consume any future coupons of these assets but only of his inventory left, $(I_t^i - A_t^i)$. With probability σ the primary dealer has no match and consumes the coupons of his inventory of assets. The inventory of off-the-run assets, I_t^f, depends on whether the primary dealer had a match in the previous period and, if so, how much was traded. The inventory is therefore endogenous. The primary dealer is also endowed with the newly issued assets I_t^2 and I_t^n. As they are just issued, their inventory is of size $\mathcal{I}$. The three inventory quantities are the state variables.

The amounts A_t^i sold and the price p_t^i are determined by a bargaining problem between the primary dealer and the seller. I assume that the primary dealer has full bargaining power.[30] This means that he sets the price just as high such that the seller is indifferent between delivery and non-delivery, i.e. $\left(q_{t-1}^i - p_t^i + \omega_{t-1}^i\right) a_t^i = q_{t-1}^i a_t^i$. The price therefore equals the collateral value ω_{t-1}^i:

$$p_t^i = \omega_{t-1}^i \qquad \forall\, i \text{ and } t. \tag{3.1}$$

The primary dealer maximises his trade surplus. The Lagrange function to his maximisation problem in each period t is given by:[31]

$$\begin{aligned} \mathcal{L}(\{A_t^i, p_t^i, \lambda_t^i, \widetilde{\lambda}_t^i\}_i) = & \sum_i \left(p_t^i A_t^i - \kappa(A_t^i) - \delta A_t^i - \beta\delta A_t^2\right) + \lambda_t^i \left[I_t^i - A_t^i\right] + \widetilde{\lambda}_t^i \left[a_t^i - A_t^i\right] \\ & + \beta(1-\sigma)\lambda_{t+1}^f \left[\mathcal{I} - A_t^2\right]. \end{aligned} \tag{3.2}$$

The trade surplus is given by the income generated, $p_t^i A_t^i$, minus the costs $\kappa(A_t^i)$ and the opportunity costs in terms of coupons, $-\delta A_t^i - \beta\delta A_t^2$. Note that for the asset maturing in two periods, the primary dealer takes into account that if he sells the asset today, he not only forgoes the coupon today but also tomorrow. For each type of asset, the primary dealer faces an inventory constraint, $A_t^i \leq I_t^i$. He cannot sell more than what he has. In addition, he cannot

[30]Given that the seller already has skin in the game, due to having sold a certain number of contracts, this is a reasonable assumption.

[31]If the primary dealer does not sell the two-period assets today, then the assets remain in his inventory in the next period if he has no match, or if he has a match, I assume that a part is sold and the rest remains in his inventory as well. This means that his inventory constraint on the off-the-run assets is not binding if he still has the full inventory available in the next period. I will show later why I assume that this holds in equilibrium and that I can always find equilibria where it does. The surplus can be written as $p_t^i A_t^i - \kappa(A_t^i) - \delta A_t^i - \beta\left[\sigma\delta A_t^2 + (1-\sigma)\left(p_{t+1}^f A_{t+1}^f - \kappa(A_{t+1}^f) + \delta\left(A_t^2 - A_{t+1}^f\right)\right)\right]$.

sell more than what the seller is willing to buy of each type of asset given by a_t^i. Therefore the following constraint needs to hold: $A_t^i \leq a_t^i$. The primary dealer makes a take-it-or-leave-it offer to the seller subject to his delivery constraints. He sets the price as such that the seller is just as well off with the purchase as without it. In both cases the seller receives the contract price q_{t-1}^i. If he delivers he has to buy the asset at price p_t^i but he can keep his collateral ω_{t-1}^i.

The first order conditions are:

$$\begin{aligned} p_t^2 &= \kappa'(A_t^2) + \delta + \beta\delta + \lambda_t^2 + \widetilde{\lambda}_t^2 + \beta(1-\sigma)\lambda_{t+1}^f \\ p_t^f &= \kappa'(A_t^f) + \delta + \lambda_t^f + \widetilde{\lambda}_t^f \\ p_t^n &= \kappa'(A_t^n) + \delta + \lambda_t^n + \widetilde{\lambda}_t^n. \end{aligned} \tag{3.3}$$

The prices equal the marginal costs faced when selling the assets and take into account potentially binding constraints. λ_t^i is the Lagrange multiplier of the inventory constraint $A_t^i \leq I_t^i$. $\widetilde{\lambda}_t^i$ is the Lagrange multiplier of the demand constraint $A_t^i \leq a_t^i$.

The complementary slackness conditions are:

$$\begin{aligned} \lambda_t^i(I_t^i - A_t^i) &= 0 \qquad \forall\, i \text{ and } t \\ \widetilde{\lambda}_t^i(a_t^i - A_t^i) &= 0 \qquad \forall\, i \text{ and } t. \end{aligned}$$

The inventories of the newly issued assets equal $\mathcal{I}$, i.e. $I_t^2 = I_t^n = \mathcal{I}$. The inventory of off-the-run assets I_t^f can take two values. The primary dealers who had a match the period beforehand have an inventory of $I_t^{f,h} \equiv I_t^f = \mathcal{I}$. The one who did not have a match have an inventory of $I_t^{f,l} \equiv I_t^f = \mathcal{I} - A_{t-1}^2$. The letter h stands for high and the letter l for low, corresponding to the higher and lower inventories, respectively. Given the probability of having no match is σ, by the law of large numbers a share σ of primary dealers has an inventory of $I_t^{f,h}$ and a share $(1-\sigma)$ has an inventory of $I_t^{f,l}$. I denote the Lagrange multiplier of the inventory constraint of the high inventory group $\lambda_t^{f,h}$ and that of the low inventory group $\lambda_t^{f,l}$. I denote the sold off-the-run assets of the high inventory group $A_t^{f,h}$ and those of the other group $A_t^{f,l}$. The other Lagrange multipliers are the same for both groups. In the following I assume that $\lambda_t^n = \lambda_t^2 = \lambda_t^{f,h} = 0$. The equilibrium where this holds is our main equilibrium of interest. In section 3.12.1 I discuss other equilibria. In these other equilibria, the main dynamics are the same. They are extreme cases of the main equilibrium.

3.4.2 Buyer

The buyer and seller anticipate which of the primary dealers' inventory constraints are non-binding and which are potentially binding when trading. They do this because they know the distribution over inventories. As mentioned above, I assume that $\lambda_t^n = \lambda_t^2 = \lambda_t^{f,h} = 0$. This means that primary dealers are unconstrained in selling assets if they still have the full amount of the assets in their inventory. Given this assumption, I have to distinguish two cases: $\lambda_t^{f,l} > 0$ and $\lambda_t^{f,l} = 0$. A fraction $(1-\sigma)$ of primary dealers has already sold a part of its stock of assets in the previous period and if $\lambda_t^{f,l} > 0$ they face a demand for assets today that will exhaust the remaining stock. If $\lambda_t^{f,l} = 0$ the demand is lower than the remaining stock.

The buyer's value function at the beginning of the contract market is

$$\begin{aligned} V^b(a_{t-1}^2) = \max_{\{a_t^i\}_i} & \sum_i -\beta q_{t-1}^i a_t^i + \beta(\delta+g)\left(\sum_i a_t^i + a_{t-1}^2\right) \\ & + \beta(1-\sigma)\left[\omega_{t-1}^f - (\delta+g)\right](a_t^f - I_t^{f,l})\mathbb{I}_{\lambda_{t+1}^{f,l}>0} + \beta V^b(a_t^2). \end{aligned}$$

The buyer's state variable is a_{t-1}^2. These are the assets he bought last period and which did not mature yet. For each asset, the buyer chooses how many he wants to buy from the seller. For each asset he wants to buy, he must build up the payment $q_{t-1}^i a_t^i$ in the form of settlement goods the next period at settlement. The assets are delivered in the next period at settlement, and the buyer receives utility $(\delta+g)$ from each asset he holds. If I am in the case where $\lambda_t^{f,l} > 0$, then with a probability $(1-\sigma)$ his seller encounters a primary dealer who is constrained in his inventory of off-the-run assets and only $I_t^{f,l}$ instead of a_t^f assets are delivered. For the amount of assets for which there is a settlement failure $(a_t^f - I_t^{f,l})$, he receives the collateral $\omega_{t-1}^f a_t^f$.

The first order conditions are

$$\begin{aligned} & q_{t-1}^2 \geq (1+\beta)(\delta+g) \\ & q_{t-1}^n \geq (\delta+g) \\ & q_{t-1}^f \geq (\delta+g) + (1-\sigma)\left[\omega_{t-1}^f - (\delta+g)\right]\mathbb{I}_{\lambda_{t+1}^{f,l}>0}. \end{aligned} \tag{3.4}$$

The prices are greater or equal the discounted marginal utilities. The price of the two-period on-the-run asset is twice the price of the one-period on-the-run asset before adjusting for discounting. If I am in the case where settlement fails can occur, i.e. $\lambda_t^{f,l} > 0$, then the price of the off-the-run asset also reflects the risk of a settlement failure.

3.4.3 Seller

The seller's value function at the beginning of the contract market is[32]

$$
\begin{aligned}
V^s = \max_{\{a^i_t\}_i} \sum_i - \omega^i_{t-1} a^i_t + \beta(1+r_t) \sum_i (q^i_{t-1} - p^i_t + \omega^i_{t-1}) a^i_t \\
+ \beta(1+r_t)(1-\sigma)(p^f_t - \omega^f_{t-1})(a^f_t - I^{f,l}_t)\mathbb{I}_{\lambda^{f,l}_{t+1}>0} + \beta V^s.
\end{aligned}
$$

The seller chooses the optimal number of contracts to sell to the buyer. That is, he chooses for each asset the amount he is willing to deliver in the next period. He has to build up collateral $\omega^i_{t-1} a^i_t$ in the form of settlement goods to support a contract. ω^i_t is taken as given. In the next period, the seller goes to the spot market and buys the assets. For each asset he can deliver, he receives the price q^i_{t-1} from the buyer, he pays the spot market price p^i_t to the primary dealer, and he can keep his accumulated collateral (all in the form of settlement goods).[33] Any remaining funds after the trade, he can deposit in the central bank facility and receive an interest rate r on them.[34] If I am in the case where $\lambda^{f,l}_t > 0$, then with probability $(1-\sigma)$ he is matched with a primary dealer who is constrained and can only deliver $I^{f,l}_t$ instead of a^f_t. For this amount of non-deliverable assets $(a^f_t - I^{f,l}_t)$, the seller's collateral is seized and given to the buyer. The seller does not buy this amount on the spot market and therefore does not have to pay the spot market price.

The first order conditions are

$$
\begin{aligned}
&\omega^2_{t-1} \geq \beta(1+r_t)(q^2_{t-1} - p^2_t + \omega^2_{t-1}) \\
&\omega^f_{t-1} \geq \beta(1+r_t)(q^f_{t-1} - p^f_t + \omega^f_{t-1}) + \beta(1+r_t)(1-\sigma)(p^f_t - \omega^f_{t-1})\mathbb{I}_{\lambda^{f,l}_{t+1}>0} \qquad (3.5) \\
&\omega^n_{t-1} \geq \beta(1+r_t)(q^n_{t-1} - p^n_t + \omega^n_{t-1}).
\end{aligned}
$$

The collateral value that has to be built up today is greater or equal to the contract price he receives tomorrow and the value of the collateral he can keep minus the spot price he has to pay to acquire the asset. If $\lambda^{f,l}_t > 0$ non-delivery occurs with probability $(1-\sigma)$ and in this case he does not have to pay the spot price but he cannot keep the collateral.

[32]The seller never buys assets for himself. The reason is that there are negative gains from trade because the primary dealer and the seller value the asset the same but if they were to trade, they would face the adjustment cost.

[33]The seller delivers an asset if $p^i_t \leq \omega^i_{t-1}$, which is the case in equilibrium for all i and t. This means that the value of the collateral seized in case of non-delivery must be as high as the value of the assets he buys on the spot market. Since this constraint is always satisfied (see section 3.4.1), it is not added to the maximisation problem.

[34]In equilibrium, the funds are not negative.

3.4.4 Equilibrium and premium definition

Before I define the equilibrium I make two assumptions.

First, I assume that if the buyer and the seller are indifferent to buying more or less assets (after accounting for the probability of a settlement failure), I assume that they are willing to buy the maximum amount of assets that is profitable and that the primary dealer can sell. This implies that in each equilibrium

$$\begin{aligned} a_t^2 &= A_t^2 \\ a_t^n &= A_t^n \\ a_t^f &= A_t^{f,h}. \end{aligned} \tag{3.6}$$

Note that in any equilibrium, as soon as an inventory constraint starts to bind, so does the corresponding constraint on the contracts. Also, if one is slack, the other is slack. The only exception is $\widetilde{\lambda}_t^{f,l}$, which can be zero even if $\lambda_t^{f,l} > 0$, but not vice versa. Therefore, as I concentrate on equilibria where $\lambda_t^n = \lambda_t^2 = \lambda_t^{f,h} = 0$, then also $\widetilde{\lambda}_t^n = \widetilde{\lambda}_t^2 = \widetilde{\lambda}_t^{f,h} = 0$.

Second, I assume in the following that due to perfect competition and market regulation the contract price and the collateral values adjust in equilibrium such that the first order conditions of the buyer and the seller hold with equality. This means that there is a non-zero finite amount of contracts sold in all assets, $a_t^i \in (0, \infty)\ \forall i$.

Definition 1 (**Equilibrium**). *An equilibrium consists of*

a) the contract and spot prices of all assets ($q_{t-1}^i\ \forall i$ and $p_t^i\ \forall i$),

b) the assets contracted ($a_t^i\ \forall i$) and sold ($A_t^2, A_t^n, A_t^{f,h}$, and $A_t^{f,l}$),

c) the collateral values ($w_{t-1}^i\ \forall i$)

and the primary dealer, the buyer, and the seller behave optimally given contract prices q_{t-1}^i and collateral values w_{t-1}^i ((3.3), (3.4), (3.5)) and the delivery constraints (3.1) and market equations (3.6) are satisfied.

Next, I define the on-the-run premium.

Definition 2 (**On-the-run premium**). *The on-the-run premium is defined as $\Delta_t \equiv p_t^n - p_t^f$.*

As mentioned above, the on-the-run and off-the-run assets have the same cash flow to maturity and mature on the same day. The only difference is their issuance date. To measure the on-the-run premium I compare these two assets. A positive (negative) premium implies that the yield to maturity of the on-the-run asset is lower (higher) than that of the off-the-run asset.

3.5 Existence

This section focuses on the case where $\lambda_t^2 = \lambda_t^n = \lambda_t^{f,h} = 0$. This means that primary dealers are unconstrained if they still have the full stock of assets available, i.e. $\mathcal{I} > max(A_t^2, A_t^n, A_t^{f,h})$. As argued in the previous section 3.4.4 $\widetilde{\lambda}_t^n = \widetilde{\lambda}_t^2 = \widetilde{\lambda}_t^{f,h} = 0$ holds as well.

I argue that in an equilibrium where there is a non-zero premium, it must be the case that $\lambda_t^{f,l} > 0$ and $\widetilde{\lambda}_t^{f,l} = 0$. Therefore in this equilibrium $\mathcal{I} \in (max(A_t^2, A_t^n, A_t^{f,h}),\ A_{t-1}^2 + A_t^{f,h})$.[35] This means that primary dealers who sold some of their inventory in the previous period and can sell again today, are constrained. All other primary dealers are not constrained.

Proposition 1. *In an equilibrium where $\mathcal{I} > max(A_t^2, A_t^n,\ A_t^{f,h})$, a necessary condition for the on-the-run premium to be non-zero is $\mathcal{I} < A_{t-1}^2 + A_t^{f,h}$.*

Proof. See appendix 3.12.2.1. □

I need constrained primary dealers for an equilibrium with a positive premium, because this gives rise to settlement fails. The fails imply the premium (see section 3.6). Without settlement fails, both assets are priced the same, since they are perfect substitutes in this case.[36] Therefore, our equilibrium candidate is the equilibrium where $\lambda_t^{f,l} > 0$.

Proposition 2. *A necessary and sufficient condition for the existence of the equilibrium with $\mathcal{I} \in (max(A_t^2, A_t^n,\ A_t^{f,h}), A_{t-1}^2 + A_t^{f,h})$ is $\sigma > \frac{1-\beta(1+r_t)}{\beta(1+r_t)}\frac{\delta}{g}$, which implies positive trade in all assets. I can always find issuance sizes $\mathcal{I}$ where this equilibrium exists.*

Proof. See appendix 3.12.2.2. □

For positive demand in off-the-run assets, the probability of a settlement fail, $(1-\sigma)$, cannot be too high. Therefore the condition. Given the condition of the proposition is satisfied, I show in the proof that I can always find an $\mathcal{I}$ where our equilibrium of interest exists, i.e. $\mathcal{I} \in (max(A_t^2, A_t^n,\ A_t^{f,h}), A_{t-1}^2 + A_t^{f,h})$. I call this equilibrium from now on "premium equilibrium".

Definition 3 (**Premium equilibrium**). *The premium equilibrium is the equilibrium where $\lambda_t^{f,l} > 0$ and all other Lagrange multipliers are zero.*

I will restrict the further analysis to the premium equilibrium. In the appendix 3.12.1, I discuss other equilibria. The dynamics and intuition are the same as in the premium equilibrium.

[35] Note that I could also add the knife-edge case where $\mathcal{I} = max(A^2, A^n, A^{f,h})$ and $\lambda_t^2 = \lambda_t^n = \lambda_t^{f,h} = 0$. To make the notation easier, I omit it.

[36] I use a linear utility function, but this is true for any utility function where the assets are perfect substitutes, i.e., only the sum of the two assets matters.

I summarise the main result of this section as follows: For an equilibrium with a premium, I need that some primary dealers are inventory constrained. The equilibrium always exists if the probability to find the off-the-run assets is high enough such that buyers and sellers want to trade it.

3.6 Premium equilibrium

In this section I analyse and discuss the premium equilibrium. To motivate my theory, I first show the scatter plot between the 10 year on-the-run premium and the net outright positions of primary dealers in 10 year Treasury bonds in figure 3.6.[37] I observe that there is a negative correlation between the on-the-run premium and the net outright positions of the primary dealers with tight 95% confidence bands. In my theory primary dealer inventory risk will be key to explain the on-the-run premium.

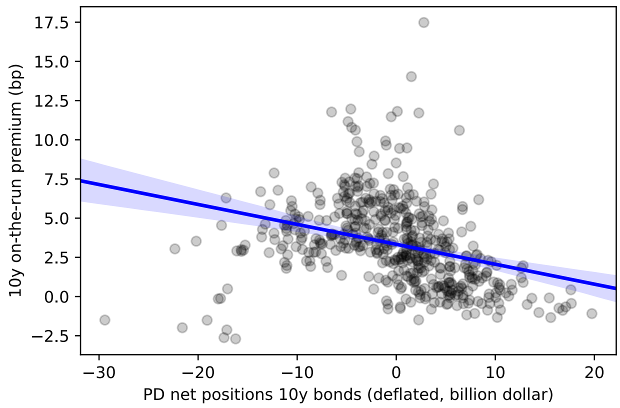


Figure 3.6: The correlation between the premium and the net positions

3.6.1 Graphical example with the two-period asset

To discuss the relevant dynamics in the premium equilibrium, I illustrate the life cycle of two two-period assets issued in period t in the figure 3.7 below (dark blue dots).

In period $t-1$ the buyer buys a contract from the seller. In period t the assets are issued and on-the-run. Every primary dealer receives one asset in his inventory. I restrict here the

[37] The data sources for the on-the-run premium are the same as in figures 3.2a. The net positions of primary dealers can be downloaded from the FED's Primary Dealer Statistics, https://www.newyorkfed.org/markets/counterparties/primary-dealers-statistics. The frequency is weekly. The data are deflated using the "Consumer Price Index for All Urban Consumers: All Items in U.S. City Average", which can be downloaded from FRED. I set the index to 1 when the series starts in 2010.

inventory to one asset for illustrative purposes. On the spot market the seller is matched with one of the two primary dealers, buys the asset and delivers it to the buyer against a payment in the form of settlement good (not illustrated).[38] The other primary dealer was not matched with a seller and keeps his asset in his inventory. After the spot market, the contract market takes place in the second subperiod. The buyer buys again one contract promising the delivery of this asset. In period $t+1$ the assets are off-the-run (as new assets are issued).[39] On the OTC spot market, the seller is matched with the primary dealer who was able to sell his asset already the period beforehand. A settlement fail occurs. Nevertheless it is optimal for the buyer to initially buy one contract. He takes the probability of a settlement fail into account when taking his decision.

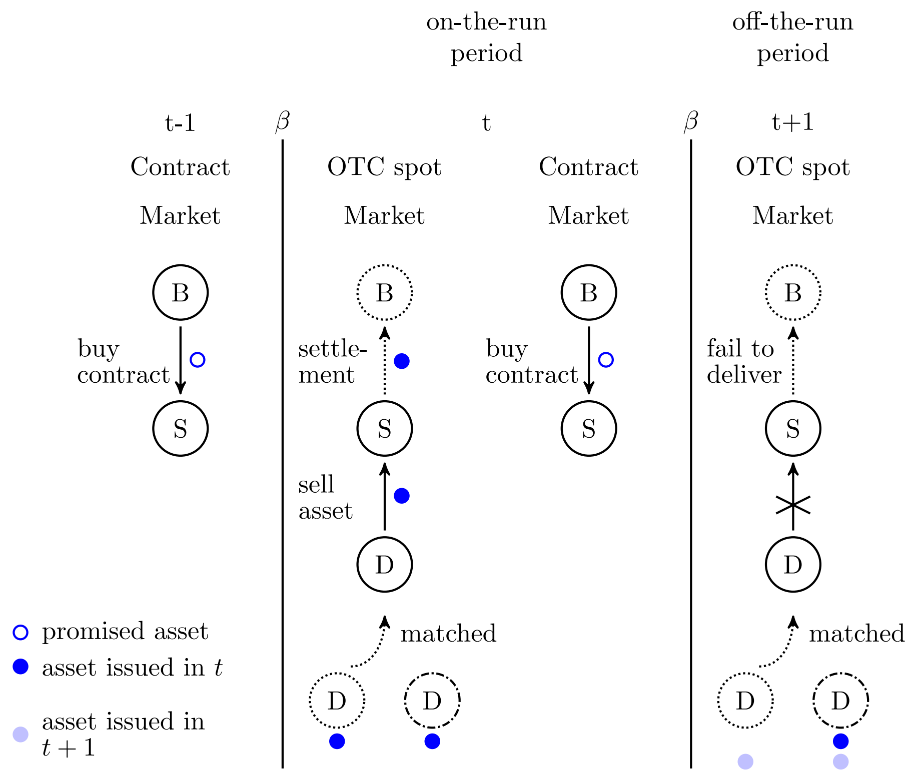


Figure 3.7: Dynamics

The figure shows only one kind of asset, the two-period asset. In the model there is also the one-period on-the-run asset. It's easy to see that settlement fails occur for off-the-run assets, but not for the on-the-run assets. Inventories do not differ for on-the-run assets because they are less long in the market. As I show in the next subsection, in equilibrium there is a premium

[38] The asset stays in the portfolio of the buyer until it matures in $t+1$.

[39] In my model the assets are always off-the-run after one period because new assets are issued (here depicted by the light blue dots). For simplicity the buyer does not buy any contract promising the delivery of these new assets in this example.

for on-the-run assets because they do not fail to settle compared to off-the-run assets where the probability of failure is priced in.

To sum up, the time since issuance is the only feature that distinguishes the two assets. And it is precisely this difference, combined with OTC market frictions and delivery constraints, that leads to settlement fails and hence the premium.[40] Finally, it is worth pointing out that off-the-run assets are scarcer (see figure). However, it is uncertainty, not scarcity per se, that causes the premium. A scarce asset that could be bought without uncertainty would not lead to fails and the premium.

3.6.2 Equilibrium prices, quantities, premium, and fails

In the premium equilibrium, the following equations for prices and quantities hold:

$$\begin{aligned}
p_t^n &= \beta(1+r_t)(\delta+g) \\
p_t^f &= \frac{\sigma}{1-(1-\sigma)\beta(1+r_t)}\beta(1+r_t)(\delta+g) \\
p_t^2 &= (1+\beta)\beta(1+r_t)(\delta+g)
\end{aligned}$$

and

$$\begin{aligned}
A_t^n &= \kappa'^{-1}\left[\beta(1+r_t)(\delta+g)-\delta\right] \\
A_t^{f,h} &= \kappa'^{-1}\left[\frac{\sigma}{1-(1-\sigma)(1+r_t)\beta}\beta(1+r_t)(\delta+g)-\delta\right] \\
A_t^{f,l} &= \mathcal{I}-A_{t-1}^2 \\
\kappa'(A_t^2) &= W+\beta(1-\sigma)\kappa'(\mathcal{I}-A_t^2)
\end{aligned}$$

where $W \equiv \left[(1+\beta)-\beta(1-\sigma)\frac{\sigma}{1-(1-\sigma)\beta(1+r_t)}\right]\beta(1+r_t)(\delta+g)-(1+\beta\sigma)\delta.$

I can observe from the equations that the spot prices directly depend on the buyer's "valuation" of the assets $(\delta+g)$. The on-the-run price of the two-period asset is $(1+\beta)$ times the on-the-run price of the one-period asset, because the buyer receives twice utility from it. All of my results with respect to the premium would also hold if I would compare the two-period off-the-run asset to the two-period on-the-run asset and abstract from the cash flow in the second period to make them equal in terms of cash flow.

[40] Without settlement fails, both assets would be priced only according to the marginal utility that the coupons (incl. additional utility g) give to the buyer. This is true not only for any linear utility function like the one used here, but also for any non-additively separable non-linear function. Once a buyer has obtained the assets, there is no reason why the on- and off-the-run assets should not be substitutes, given that they have the same coupons and are held to maturity.

The quantity of the two-period asset maturing in two periods and the off-the-run asset traded by constrained dealers depends on the issue size $\mathcal{I}$. The reason for the first mentioned fact is that when these two-period assets are sold, it is taken into account that less can be sold tomorrow due to the binding inventory constraint. This binding inventory constraint is then also the reason why the amount of off-the-run assets traded by constrained primary dealers depends on $\mathcal{I}$.

Lastly I derive the on-the-run premium. I deduct p_t^f from p_t^n. This gives rise to the premium according to the following proposition:

Proposition 3. *The on-the-run premium is given by:* $\Delta = \left[1 - \frac{\sigma}{1-(1-\sigma)\beta(1+r_t)}\right]\beta(1+r_t)(\delta + g) > 0$. *If* $\sigma \to 1$, *then* $\Delta \to 0$.

The on-the-run premium depends on the buyer's asset "valuation" $(\delta + g)$, the probability to find the off-the-run assets σ, and $\beta(1+r_t)$ which is the discount factor cost of the seller for binding collateral. If assets would be found with certainty in the second period, i.e. $\sigma \to 1$, then the premium vanishes.[41]

It is important to point out that key for a positive on-the-run premium to arise is

$$I_t^{f,h} \neq I_t^{f,l}.$$

In the premium equilibrium, unconstrained primary dealers have a full inventory, i.e. $I_t^{f,h} = \mathcal{I}$, and constrained ones have a reduced inventory, i.e. $I_t^{f,l} = \mathcal{I} - A_{t-1}^2$ with $A_{t-1}^2 > 0$. Therefore, $I_t^{f,h} > I_t^{f,l}$. It is this difference in inventories that leads to the uncertainty that implies settlement fails. The fails themselves imply the premium. Compared to Vayanos and Weill (2008) the premium is always on the on-the-run asset.

Lastly, I define the settlement failure rate of asset i as the value of assets i involved in a fail divided by the overall amount of assets promised to be delivered:

$$f_t^i \equiv \frac{p_t^i \mathbb{I}_{a_t^i > I_t^i} \mathbb{P}_t^i (a_t^i - I_t^i)}{p_t^i a_t^i}$$

where $\mathbb{P}_t^i$ is the failure probability. As $\lambda_t^{f,h} = 0$ it follows that $f_t^n = 0$ and as $\lambda_t^{f,l} > 0$ it follows that $f_t^f = \frac{(1-\sigma)p_t^f(a_t^f - I_t^f)}{p_t^f a_t^f} = (1-\sigma)\left(1 - \frac{\mathcal{I} - a_t^2}{a_t^f}\right)$.

Our equilibrium is consistent with the four stylised facts described in the introduction. First, on-the-run assets are more expensive than off-the-run assets (positive on-the-run premium).

[41]The premium also vanishes if $\sigma = 0$ but then I am not anymore in the premium equilibrium and the formula above does not hold.

Second, they trade in larger volumes. Third, settlement fails occur and lastly, off-the-run assets fail to settle more often. I show this in the following proposition:

Proposition 4. *In the premium equilibrium*

$$
\begin{aligned}
&p_t^n > p_t^f \\
&A_t^n > \sigma A_t^{f,h} + (1-\sigma) A_t^{f,l} \\
&f_t^f > 0 \\
&f_t^f > f_t^n .
\end{aligned}
$$

Proof. See appendix 3.12.2.3. □

In equilibrium not only $p_t^n > p_t^f$ but also $A_t^n > \sigma A_t^{f,h} + (1-\sigma) A_t^{f,l}$, i.e. on-the-run assets not only have a higher price but are also traded in larger quantities. Both equilibrium results can be explained by the fact that off-the-run assets are less attractive because they fail to settle more often. It is contrary to common intuition that a scarcer asset has the lower price, but this observation is consistent with what is observed in the market (U.S. Department of the Treasury (2022)).

I summarise the above discussion as follows: The on-the-run premium is due to differences in inventories of off-the-run assets as they are longer in the market. The reason is as follows: Some of the off-the-run assets are locked up in buy-and-hold portfolios because they have already been sold during their on-the-run period. Since not all primary dealers faced the same demand during the on-the-run period due to the OTC market structure, there are differences in their inventories at the start of the off-the-run period. This implies uncertainty about the amount of assets available in an upcoming match with them in the OTC market. This leads to a higher frequency of settlement fails for off-the-run assets, as contracts promising their delivery cannot always be fulfilled. There is a preference for on-the-run assets because their settlement is not risky. Compared to off-the-run assets they are safe in this aspect. This implies that they carry a premium, i.e. are more expensive on the spot market, and trade in larger quantities than off-the-run assets.

3.7 On-the-run premium, settlement fails, and inventories in the data

Next, I analyse whether my theory is consistent with the empirical evidence. From the theory I derive two hypotheses, which I test:

Hypothesis 1. *A higher failure rate leads to a higher premium.*

Hypothesis 2. *Lower primary dealer inventories lead to a higher failure rate.*

The first hypothesis follows directly from proposition 3. There I show that if the probability of finding the assets, σ, goes to 1 (no fails occur), then the premium vanishes. Otherwise, there are settlement fails that lead to the premium. Therefore, in a first step, I regress on-the-run premium data on the failure rate of off-the-run assets. My theory predicts a negative coefficient.

Second, I conjecture that lower primary dealer inventories lead to settlement fails. To test the second hypothesis, I regress the failure rate of off-the-run assets on the net outright positions in Treasuries of the primary dealers. Lower net outright positions are expected to capture inventory uncertainty. Given my theoretic results, I should observe a negative coefficient.

I first examine the relationship between the failure rate and the on-the-run premium. For the outcome variable in the first regression (see table 3.1), I use data of the 1, 2, 3, 5, 7, 10, 20, and 30 year on-the-run premia. Each maturity, denoted by m and day, denoted by t, in the sample is a separate observation. The explanatory variable, the failure rate of off-the-run Treasuries, is not available for different maturities. The explanatory variable is therefore the same for each on-the-run premium maturity. I run three regressions. In the first, I only regress on the failure rate. In the second, I add other control variables. The additional control variables are the logarithm of the VIX, the 10 year - 2 year yield spread, and the general collateral financing repo rate. In the third regression, I add additionally crisis and maturity fixed effects. The crisis fixed effects capture the months October 2014, September 2019, and March 2020. In these months there was a crisis in the Treasury market (see U.S. Department of the Treasury et al. (2015), Anbil et al. (2020), and Schrimpf et al. (2020)). As we use weekly data, the crisis dummies are equal to one in each week within these months. The maturity fixed effects capture each maturity m of the on-the-run premia. The regression equation (3) is given by

$$\begin{aligned}\text{On-the-run premium}_{t,m} = {} & \alpha_m + \beta_1 \text{ Failure rate off-the-run Treasuries}_t + \beta_2\, ln(\text{VIX}_t) \\ & + \beta_3 \text{ 10 year - 2 year yield spread}_t \\ & + \beta_4 \text{ General collateral financing repo rate}_t \\ & + \beta_6\, \mathbb{I}_{10/14,t} + \beta_7\, \mathbb{I}_{09/19,t} + \beta_8\, \mathbb{I}_{03/20,t} + u_t.\end{aligned}$$

The results are shown in table 3.1. As expected, the coefficient of the failure rate is positive. It is significant at the 5% level (with and without additional control variables and fixed effects). If I do not control for any other variables, then an increase in the failure rate by 1 percentage point, increases the premium by 2.2 basis points, which is about half a standard deviation. The size of the effect is almost the same if I add the additional control variables.

	(1) Premium	(2) Premium	(3) Premium
Failure rate off-the-run Treasuries	2.2483***	2.4206***	2.3150***
	(0.630)	(0.563)	(0.503)
lnVIX		1.1706***	1.1653 ***
		(0.570)	(0.492)
10 year - 2 year yield spread		0.2605	0.1965
		(0.364)	(0.346)
General collateral financing repo rate		0.0128	-0.0364
		(0.314)	(0.287)
Constant	1.1459***	-2.5491	-2.0929
	(0.412)	(2.054)	(1.878)
Crisis and maturity fixed effects	No	No	Yes
No. Observations:	3760	3760	3760
R-squared:	0.017	0.025	0.219
Adj. R-squared:	0.017	0.024	0.216

Notes: I use Newey-West standard errors with 6 lags. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level. The sample period is April 2013-2022 and the frequency is weekly.
The data source on the premia (in basis points) is the same as in figure 3.2a. The failure rate data (in percent) is the same as in figure 3.3a. The VIX data are taken from FRED. The 10 year and 2 year yields (par yields of seasoned bonds in percent) are taken from the FED yield curve, which is an updated version of the original Gürkaynak-Sack-Wright curve (Gürkaynak et al. (2010)), https://www.federalreserve.gov/data/tips-yield-curve-and-inflation-compensation.htm. The general collateral financing repo rate (in percent) is provided by the Depository Trust and Clearing Corporation (DTCC) and can be downloaded here: https://www.dtcc.com/charts/dtcc-gcf-repo-index. The crisis fixed effects capture the months October 2014, September 2019, and March 2020 and the maturity fixed effects each maturity of the on-the-run premia (1y, 2y, 3y, 5y, 7y, 10y, 20y, and 30y).

Table 3.1: Premium on off-the-run failure rate regression

Next, I examine the relationship between the failure rate and dealer inventories. I regress the failure rate of off-the-run Treasuries on the net outright positions of primary dealers in Treasuries of different maturity baskets (see table 3.2). The maturity of the baskets is denoted by k. Specifically, I have the following baskets for the net outright positions in Treasuries: below or equal 3 years, above 3 years to 6 years, above 6 years to 11 years, above 11 years. Each basket and day is a separate observation. I run the regression first without and then with additional control variables. The control variables are the same as in the previous regression (see table 3.1). In a third additional regression, I add crisis fixed effects also the same ways as before. The regression equation (3) is given by

$$\begin{aligned}\text{Off-the-run failure rate}_t = \alpha &+ \beta_1 \text{ Primary dealer net positions}_{t,k} + \beta_2\, ln(\text{VIX}_t)\\ &+ \beta_3 \text{ 10 year - 2 year yield spread}_t\\ &+ \beta_4 \text{ General collateral financing repo rate}_t\\ &+ \beta_6\, \mathbb{I}_{10/14,t} + \beta_7\, \mathbb{I}_{09/19,t} + \beta_8\, \mathbb{I}_{03/20,t} + u_t.\end{aligned}$$

	(1) Off-the-run failure rate	(2) Off-the-run failure rate	(3) Off-the-run failure rate
PD net positions (deflated)	-0.0012***	-0.0031***	-0.0030***
	(0.001)	(0.000)	(0.000)
lnVIX		-0.0769***	-0.0786***
		(0.033)	(0.033)
10 year - 2 year yield spread		-0.1559***	-0.1637***
		(0.019)	(0.019)
General collateral financing repo rate		-0.0256**	-0.0324***
		(0.014)	(0.014)
Constant	0.7208***	1.1508***	1.1673***
	(0.015)	(0.111)	(0.110)
Crisis fixed effects	No	No	Yes
No. Observations:	1912	1880	1880
R-squared:	0.009	0.132	0.145
Adj. R-squared:	0.008	0.130	0.142

I use Newey-West standard errors with 6 lags. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level. The sample period is April 2013-2022 and the frequency is weekly.
The data source for the net outright positions (in billion US dollars) is the same as in figure 3.6. The data are also deflated in the same way (with the index set to 1 when the series starts in April 2013). The failure rate data (in percent) is the same as in figure 3.3a. The VIX data are taken from FRED. The 10 year and 2 year yields (par yields of seasoned bonds in percent) are taken from the FED yield curve, which is an updated version of the original Gürkaynak-Sack-Wright curve (Gürkaynak et al. (2010)), https://www.federalreserve.gov/data/tips-yield-curve-and-inflation-compensation.htm. The general collateral financing repo rate (in percent) is provided by the Depository Trust and Clearing Corporation (DTCC) and can be downloaded here: https://www.dtcc.com/charts/dtcc-gcf-repo-index. The crisis fixed effects capture the months October 2014, September 2019, and March 2020.

Table 3.2: Failure rate on net positions regression

The results are shown in table 3.2. As expected, the coefficient of the primary dealers net positions is negative and significant at the 5% level.[42] Adding the additional control variables and fixed effects does not change the result. As the net positions are denominated in dollars, the coefficient is best interpreted by first multiplying it by the standard deviation of the net outright positions. If the net positions increase by one standard deviation, then the failure rate

[42] The result still holds if the first difference in the primary dealers' net positions is taken.

reduces by 2, 6, resp. 5 basis points (regression 1-3).[43]

To provide further evidence consistent with my theory, I examine the volatility of the net positions of primary dealers in Treasuries.[44] I find evidence in the data consistent with my theory that off-the-run inventories are much more affected by search and matching frictions than on-the run inventories. Table 3.3 shows the volatilities. Primary dealer inventories in on-the-run Treasuries are much less volatile than inventories including all kind of Treasuries. This holds across all maturity baskets. This is consistent with my theory if I assume that there is a slight variation in σ over time, which must be the case in reality. The effect of the search and matching frictions on off-the-run inventories can then be observed. They are volatile, while on-the-run inventories are much more stable.

On-the-run	Maturity	2y	3y	5y	7y	10y	30y
	Volatility	4.1	4.0	3.5	2.7	4.1	2.3
All	Maturity	≤3y	(3y,6y]	(6y,7y]	(7y,11y]	≥11y	
	Volatility	21.2	11.7	7.6	7.0	12.3	

Table 3.3: Primary dealer net positions volatilities

3.8 Central bank facility access

How does broadening access to a central bank's reverse repo or deposit facility to any type of intermediary affect the Treasury market (in normal times)? How do prices, premia, traded quantities, fails, and profits change? This section provides answers to these questions.

In my model, the central bank facility is comparable to a deposit facility or a reverse repo facility, where I abstract from the collateral provided. The collateral part would not make the existing dynamics disappear. First, the repos from the facility are general collateral repos. For such repos the cash lender is willing to accept any collateral that falls into a broad class, and is not looking for any particular collateral (Bowman et al. (2017)). Second, even if the facility were to provide a specific Treasury sought, the Treasury would have to be returned the next day and would only temporarily ease availability.

[43]This corresponds to 0.1 (regression 1), 0.2 (regression 2), respectively, standard deviations.

[44]The data can be downloaded from the FED's Primary Dealer Statistics, https://www.newyorkfed.org/markets/counterparties/primary-dealers-statistics. The data are deflated using the "Consumer Price Index for All Urban Consumers: All Items in U.S. City Average", which can be downloaded from FRED. I set the index to 1 when the on-the-run net positions time series start in April 2013. I use data from April 2013 to the end of 2022. The frequency is weekly. To calculate the volatilities, I use the average of the data over the full time horizon.

To motivate my theory I show the scatter plot between the 10 year on-the-run premium and the reverse repo facility rate in figure 3.8.[45] I observe that there is a negative correlation between the 10 year on-the-run premium and the reverse repo facility rate with tight 95% confidence bands. As I show below, this empirical evidence is consistent with my theory.

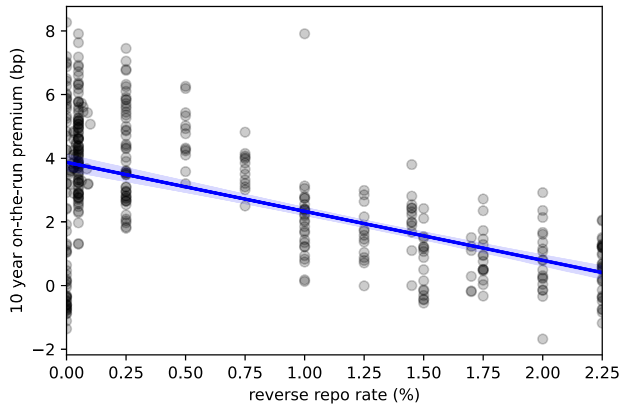


Figure 3.8: The correlation between the premium and the reverse repo facility rate

3.8.1 Homogenous sellers

First, I analyse the situation where all sellers have or gain access to the central bank facility. The facility provides liquidity in-between transactions. It can be accessed every first subperiod for one subperiod. Sellers can deposit settlement good and receive an interest rate r_t on it. Sellers do this because they have a positive net settlement good position after the trades: They received the contract price q_t^i and had to pay (in case of delivery) the lower spot price p_t^i and could keep the collateral ω_t^i. The seller represents any type of financial institution (e.g., a hedge fund or a non-primary dealer). The goal of the analysis is to find, in a general setting, the effect of the reverse repo or deposit facility rate on trading (in normal times) when any type of financial firm other than a primary dealer is given access to the facility.

I still look at the premium equilibrium. I assume a permanent unanticipated increase in the facility rate r_t at the beginning of the contract market in $t = \widetilde{t} - 1$. In addition to the situation where all sellers already had access and the facility rate increases, the increase can also represent the situation where all sellers gain access to the facility and because of the facility now face a higher interest rate than before (with no interest or a lower market rate). I assume

[45] The data source for the premium is the same as in figure 3.2a. The reverse repo facility rate data is provided by the New York FED and can be downloaded here: https://www.newyorkfed.org/markets/omo_transaction_data#rrp. I take daily averages. The time horizon is 23 September 2013-2021 Q2 and the frequency is weekly.

that there is still a positive premium after the sellers gain access. The results of the analysis can be summarised as follows:

Corollary 1. *An unanticipated and permanent increase in r_t in the second subperiod in $t = \widetilde{t}-1$*

a) increases all spot prices, $\frac{dp_t^i}{dr_{\widetilde{t}-1}} > 0 \, \forall i$ and $\forall t \geq \widetilde{t}$,

b) decreases the on-the-run premium, $\frac{d\Delta_t}{dr_{\widetilde{t}-1}} < 0 \, \forall t \geq \widetilde{t}$,

c) increases the quantities of on-the-run and off-the-run assets traded if no inventory constraint binds, $\frac{dA_t^n}{dr_{\widetilde{t}-1}} > 0$ and $\frac{dA_t^{f,h}}{dr_{\widetilde{t}-1}} > 0 \, \forall t \geq \widetilde{t}$,

d) implies that the quantities of off-the-run assets traded initially, i.e. in $t = \widetilde{t}$, stay the same and then decrease if inventory constraints bind, $\frac{dA_t^{f,l}}{dr_{\widetilde{t}-1}} = 0$ for $t = \widetilde{t}$ and $\frac{dA_t^{f,l}}{dr_{\widetilde{t}-1}} < 0$ $\forall t > \widetilde{t}$ if $\sigma > \widetilde{\sigma}$,

e) implies that overall more assets are offloaded from the inventories and end up in the portfolio of the buyer, i.e. the holder with the highest marginal asset valuation, $\frac{d(A_t^2 + A_t^n + \sigma A_t^{f,h} + (1-\sigma)A_t^{f,l})}{dr_{\widetilde{t}-1}} > 0$ for $t = \widetilde{t}$ and $\forall t > \widetilde{t}$ if $\sigma > \widetilde{\sigma}$,

f) increases the settlement failure rate of the off-the-run asset, $\frac{df_t^f}{dr_{\widetilde{t}-1}} > 0$ for $t = \widetilde{t}$ and $\forall t > \widetilde{t}$ if $\sigma > \widetilde{\sigma}$.

Proof. See appendix 3.12.2.5 □

The value $\widetilde{\sigma}$ is the value of σ above which always (but not only) $\frac{\partial A_t^2}{\partial r_{\widetilde{t}-1}} > 0 \, \forall t \geq \widetilde{t}$. I define it in appendix 3.12.2.5 and show that with a reasonable calibration it is an empirically very small value.

The intuition for this result is as follows: An increase in the facility rate increases the profitability of the trade for the seller. The facility is always available, and excess liquidity can be deposited until the seller enters the next trade. The facility is not a substitute for trading, but a complement to it. The higher the interest rate, the higher the profitability. Increased profitability leads to a positive supply shock in the contract market. This in turn implies a positive demand shock in the spot market. In equilibrium, spot prices and quantities traded by unconstrained primary dealers increase.

Off-the-run prices and quantities react more strongly than their on-the-run equivalents, and the premium falls. The reason is that the case of non-delivery for the off-the-run asset is less costly than before because the collateral values increase due to the rise in prices. This reduces the spread between the real valuation of holding the asset and the collateral. This effect additionally triggers the demand for off-the-run assets. This makes the policy particularly interesting in the context of the Treasury market crisis during the pandemic, where the market for off-the-run assets froze (Eren and Wooldridge (2021)).

The amount of off-the-run assets traded by constrained dealers remains the same in the first period after the rate hike, because inventories are determined from the previous period. Later, it decreases because the inventories of the constrained dealers are smaller because more assets have already been sold during their on-the-run period (if search frictions are above a small minimum value of $\widetilde{\sigma}$). This is also the main reason for the observed increase in the settlement failure rate of off-the-run assets. The assets are not available later and more fails occur. Nevertheless, due to the interest rate increase, more assets end up in the portfolio of the buyer, the agent with the highest marginal asset valuation.

Putting all the above results in a broader context, I conclude that any kind of policy that lowers the costs of trade and intermediation can trigger the effects described. Access is one possibility.

The next result is about who benefits from an increase in the facility rate r_t. I look at the impact on the lifetime values of the buyer, the seller, and the primary dealer. There are two groups of primary dealers: One group of primary dealers has an inventory of $\mathcal{I}$ off-the-run assets and another group has an inventory of $\mathcal{I} - A_t^2$. I take the lifetime value of both groups and average them according to their proportions in the population. To simplify the notation, I define $V_{\widetilde{t}-1}^{D,a} \equiv (1-\sigma)V_{\widetilde{t}-1}^{D}(\mathcal{I} - A_t^2) + \sigma V_{\widetilde{t}-1}^{D}(\mathcal{I})$.

Corollary 2. *An unanticipated and permanent increase in r_t in the second subperiod in $t = \widetilde{t}-1$*

a) does not affect the lifetime value of sellers, $\frac{dV_{\widetilde{t}-1}^{s}}{dr_{\widetilde{t}-1}} = 0$,

b) decreases the lifetime value of buyers, $\frac{dV_{\widetilde{t}-1}^{b}}{dr_{\widetilde{t}-1}} < 0$,

c) increases the lifetime value of primary dealers, $\frac{dV_{\widetilde{t}-1}^{D}}{dr_{\widetilde{t}-1}} > 0$ *if* $\sigma > \widetilde{\sigma}$.

Proof. See appendix 3.12.2.6. □

As pointed out above, see appendix 3.12.2.5 for the definition of $\widetilde{\sigma}$.

The intuition for the result is as follows: Competition among sellers in the contract market erodes any positive profits for them to zero. Initially increased profitability is offset by higher spot prices in equilibrium.

Primary dealers benefit from the policy by providing the assets that are in higher demand. They sell more assets, they sell them earlier, and they sell them at a higher price. In equilibrium, primary dealers sell assets until their marginal nonlinear cost equals the price. Therefore, each asset sold yields a small positive marginal surplus until the last asset is sold, where the marginal surplus equals zero. Since prices are higher in the new equilibrium, breakeven is reached at a higher quantity of assets. Primary dealers profits increase.

Buyers' utility falls. To gain intuition, I first explain how buyers' utility or benefit materialises in equilibrium. Since the price of the off-the-run asset reflects the utility of the last unit bought, it incorporates the probability that a settlement fail occurs. But the first part of the bought assets is found with certainty, the part I_t^f. The value the buyer places on these assets is therefore higher. Nevertheless, in equilibrium he pays the same price for all the assets and therefore he has a small benefit.

An increase in the facility rate increases the off-the-run price and decreases the quantity of off-the-run assets found with certainty (as the quantity of on-the-run assets traded with a two-period maturity increases). Both effects lead to a lower profit. In summary, buyers' profits decrease as the spread between the value of an uncertain unit and a certain unit decreases and there are fewer certain units. The decrease in profits or utility is an externality problem. The buyer does not consider how the purchase of two-period assets affects the availability and price of the same assets in the next period. This is in contrast to the primary dealer, who manages his inventory and takes into account that a two-period asset sold today cannot be sold in the next period.

I solved my baseline model under the assumption that the primary dealer has full bargaining power. I can relax this assumption and prove that even with positive bargaining power of the seller, in equilibrium, the seller does not make a profit in contrast to the primary dealer.

Corollary 3. *Result 2 still holds even with positive bargaining power of sellers.*

When maximising the joint surplus of both agents, the first-order conditions of the spot market problem change to:

$$
\begin{aligned}
p_t^2 &= \kappa'(A_t^2) + \delta + \beta\delta + (\omega_{t-1}^2 - p_t^2) + \lambda_t^2 + \widetilde{\lambda}_t^2 + \beta(1-\sigma)\lambda_{t+1}^f \\
p_t^f &= \kappa'(A_t^f) + \delta + (\omega_{t-1}^f - p_t^f) + \lambda_t^f + \widetilde{\lambda}_t^f \\
p_t^n &= \kappa'(A_t^n) + \delta + (\omega_{t-1}^n - p_t^n) + \lambda_t^n + \widetilde{\lambda}_t^n.
\end{aligned}
$$

The surplus is divided according to their bargaining power: $\left(\omega_{t-1}^i - p_t^i\right) A_t^i = \frac{(1-\theta)}{\theta} S_t^i$ where S_t^i is the surplus of the primary dealer when selling assets i. The first-order conditions of the seller and the buyer when making their decision about the optimal number of contracts to sell and buy do not change. This is crucial because I see from the seller's first-order conditions and his value function that he makes no profit in equilibrium. Even if in equilibrium $\omega_{t-1}^2 - p_t^2 > 0$ (due to the new pricing scheme) and the seller receives part of the positive total surplus of the OTC trade, the collateral value ω_{t-1}^i in the Walrasian contract market adjusts in such a way that he has no total profits. Otherwise, the seller would supply an infinite number of contracts or no contracts at all. I assumed that this is not the case in equilibrium, which is a reasonable assumption. Therefore, even with positive bargaining power, the seller never profits from an increase in the facility rate if he has access.

3.8.2 Heterogenous sellers

In this subsection I assume that there is a measure ξ of sellers which have access to the facility and a measure $(1-\xi)$ which does not have access. Sellers which have access are denoted by a and the ones which don't by na. The facility rate is given by r_t and the market rate by r_t^m. I assume that $r_t > r_t^m$. Agents with no access face the market rate, while the others will use the facility. Again I will look at the impact of sellers gaining access. But beforehand I present the equilibrium equations.

I again consider the premium equilibrium. I first show the prices and quantities of the newly issued assets. Analogously to section 3.6, the contract and spot prices are given by

$$\begin{aligned} q_t^n &= (\delta + g) \\ p_t^{n,na} &= \beta(1 + r_t^m)(\delta + g) \\ p_t^{n,a} &= \beta(1 + r_t)(\delta + g) \end{aligned}$$

and

$$\begin{aligned} q_t^2 &= (1+\beta)(\delta + g) \\ p_t^{2,na} &= (1+\beta)\beta(1 + r_t^m)(\delta + g) \\ p_t^{2,a} &= (1+\beta)\beta(1 + r_t)(\delta + g). \end{aligned}$$

The traded quantities are determined by the following equations:

$$\begin{aligned} A_t^{n,na} &= \kappa'^{-1}\left[\beta(1 + r_t^m)(\delta + g) - \delta\right] \\ A_t^{n,a} &= \kappa'^{-1}\left[\beta(1 + r_t)(\delta + g) - \delta\right] \end{aligned}$$

and

$$\kappa'(A_t^{2,na}) = W^{na} + \beta(1-\sigma)\kappa'(\mathcal{I} - A_t^{2,na})$$
$$\kappa'(A_t^{2,a}) = W^{a} + \beta(1-\sigma)\kappa'(\mathcal{I} - A_t^{2,a})$$

where
$W^{na} \equiv p_t^{2,na} - \beta(1-\sigma)\left[\xi p_t^{f,a} + (1-\xi)p_t^{f,na}\right] - (1+\beta\sigma)\delta$ and
$W^{a} \equiv p_t^{2,a} - \beta(1-\sigma)\left[\xi p_t^{f,a} + (1-\xi)p_t^{f,na}\right] - (1+\beta\sigma)\delta$.

Next I show the off-the-run prices and quantities. Let me denote by $\mathbb{P}^a$ the probability of a seller with access finding the off-the-run assets. Analogously, I define $\mathbb{P}^{na}$ as the probability for the sellers without access. Analogously to section 3.6, the off-the-run prices are given by

$$p_t^{f,a} = \frac{\mathbb{P}^a}{1-(1-\mathbb{P}^a)\beta(1+r_t)}\beta(1+r_t)(\delta+g)$$
$$p_t^{f,na} = \frac{\mathbb{P}^{na}}{1-(1-\mathbb{P}^{na})\beta(1+r_t^m)}\beta(1+r_t^m)(\delta+g).$$

As in section 3.5 I assume that the primary dealers with a full inventory are unconstrained. Therefore

$$A_t^{f,a,h} = \kappa'^{-1}\left[\frac{\mathbb{P}^a}{1-(1-\mathbb{P}^a)\beta(1+r_t)}\beta(1+r_t)(\delta+g) - \delta\right]$$
$$A_t^{f,na,h} = \kappa'^{-1}\left[\frac{\mathbb{P}^{na}}{1-(1-\mathbb{P}^{na})\beta(1+r_t^m)}\beta(1+r_t^m)(\delta+g) - \delta\right].$$

Lastly, I define the on-the-run premium by $\Delta \equiv [\xi p_t^{n,a} + (1-\xi)p_t^{n,na}] - [\xi p_t^{f,a} + (1-\xi)p_t^{f,na}]$. For a positive premium either $\mathbb{P}^a$ or $\mathbb{P}^{na}$ or both must be below 1. If they are below one, they either equal σ or $\sigma + (1-\sigma)(1-\xi)$ given that $(\mathcal{I} - A_{t-1}^{2,na}) > (\mathcal{I} - A_{t-1}^{2,a})$, or $A_{t-1}^{2,a} > A_{t-1}^{2,na}$.[46]

Next, I analyse the effect of sellers without access gaining access. The analysis is done analogously to section 3.8.1. As in section 3.8.1 I also assume that there is still a positive premium after all sellers have access. The results are summarised below.

Corollary 4. *If sellers without access gain access, meaning that their interest rate unanticipated permanently increases from r_t^m to r_t in the second subperiod in $t = \widetilde{t} - 1$, then*

a) all on-the-run spot prices charged by the sellers gaining access increase and all sellers charge $p_t^{j,a}$ in $t \geq \widetilde{t}$ for $j = 2$ and n,

[46]For the inventories I do not only need to distinguish if a primary dealer has met a seller the previous period or not (as before) but also if the match was with a seller with or without access. Depending on the type, more or less assets were sold and therefore the inventory differs. A mass σ of agents has a high inventory of $\mathcal{I}$ because they did not face any demand the period beforehand. A mass $(1-\sigma)\xi$ of primary dealers has sold assets to a seller with access in the previous period and they have an inventory of $(\mathcal{I} - A_{t-1}^{2,a})$. Lastly, a mass $(1-\sigma)(1-\xi)$ of sellers have an inventory of $(\mathcal{I} - A_{t-1}^{2,na})$ as they met a seller without access the previous period.

b) off-the-run spot prices charged by the sellers gaining access initially, i.e. in $t = \widetilde{t}$, increase and all sellers charge $p_{\widetilde{t}}^{f,a}$,

c) the on-the-run premium Δ_t initially, i.e. in $t = \widetilde{t}$, decreases,

d) all quantities of on-the-run assets traded by the sellers gaining access increase and all sellers trade $A_t^{j,a}$ in $t \geq \widetilde{t}$ for $j = 2$ and n,

e) quantities of off-the-run assets traded by the sellers gaining access inititally, i.e. in $t = \widetilde{t}$, increase if no inventory constraints bind and all sellers trade $A_{\widetilde{t}}^{f,a,h}$,

f) quantities of off-the-run assets traded by the sellers gaining access initially, i.e. in $t = \widetilde{t}$, stay the same if inventory constraints bind and equal $A_{\widetilde{t}-1}^{f,na,l}$.

The points b), c) and e) also hold for $t > \widetilde{t}$ if the probabilities to find the off-the-run assets do not change, otherwise the effects are ambiguous.

Proof. See appendix 3.12.2.7. □

As already pointed out in section 3.8.1, the value $\widetilde{\sigma}$ is the value of σ above which always (but not only) $\frac{\partial A_t^2}{\partial r_{\widetilde{t}-1}} > 0 \; \forall t \geq \widetilde{t}$. I define it in appendix 3.12.2.5. I also show that with a reasonable calibration it is an empirically very small value.

The result is self-explanatory. The direction of the effects on the sellers who gain access is (initially at $t = \widetilde{t}$ and also later if the probabilities of finding the off-the-run assets do not change) analogous to our baseline scenario with homogeneous sellers in section 3.8.1.

3.9 On-the-run premium and facility rate in the data

In this section I examine the relationship between the on-the run premium and the central bank facility rate. Based on corollary 1 I test the following hypothesis:

Hypothesis 3. *A higher reverse repo facility rate leads to a lower premium.*

I then regress on-the-run premia data on the reverse repo facility rate (see table 3.4).[47] My theory predicts a negative coefficient. I use data on the 1, 2, 3, 5, 7, 10, 20, and 30 year on-the-run premia. Each maturity and each day in the sample is a separate observation. The explanatory variable is the reverse repo facility rate. The first observation I use is from the

[47] I use the reverse repo facility and not any deposit facility data as the access discussion centers around this one. An increasing number of institutions already access this facility (Frost et al. (2015), Baklanova et al. (2015) and Marte (2021)).

23 September 2013, as this was the first day the facility was available on a large scale.[48] I run my regressions with and without additional control variables. The additional control variables are the deflated reverse repo facility trade amounts, the logarithm of the VIX, the 10 year - 2 year yield spread, and the general collateral financing repo rate. I also add crisis and maturity fixed effects in a third regression the same way as in regression 3.1. The regression equation (3) is given by

$$\begin{aligned}
\text{On-the-run premium}_{t,m} = \alpha_m &+ \beta_1 \text{ Reverse repo facility rate}_t \\
&+ \beta_2 \text{ Reverse repo facility trade amounts}_t + \beta_3\, ln(\text{VIX}_t) \\
&+ \beta_4 \text{ 10 year - 2 year yield spread}_t \\
&+ \beta_5 \text{ General collateral financing repo rate}_t \\
&+ \beta_6\, \mathbb{I}_{10/14,t} + \beta_7\, \mathbb{I}_{09/19,t} + \beta_8\, \mathbb{I}_{03/20,t} + u_t.
\end{aligned}$$

From table 3.4, I can see that as expected, the coefficient of the reverse repo facility rate is negative. A one percentage point increase in the reverse repo facility rate reduces the on-the-run premium by 0.01 basis points (regression 1), 2.2 basis points[49] (regression 2 and 3), respectively. The coefficient is only significant (at the 5% level) when we add the additional control variables.[50]

3.10 Life cycle model

This section presents an extension of the basic model. It shows that the model can be used as a tool to describe the life cycle of a Treasury. In addition I can explain two more stylised facts:

I When-issued Treasuries trade at a premium compared to previously issued Treasuries.

II The primary market prices are lower than the secondary market prices.

The observations are illustrated in the appendix 3.12.3. The life cycle of a Treasury can be divided into three periods as illustrated in figure 3.9: the when-issued period, the on-the-run period, and the off-the-run period. The auction takes place between the announcement and the

[48] See https://www.newyorkfed.org/markets/opolicy/operating_policy_130920.html for more information.

[49] This corresponds to around 0.7 standard deviations.

[50] The data on each reverse repo facility transaction, including the rate, the amount traded and other details, are only publicly available after two years. However, a reverse repo facility rate series can be downloaded without any time lag. When I use these data (without controlling for trade size) and extend the time horizon to 31 August 2023, the coefficient of the rate is positive in the regression without the additional control variables and still negative in the one with. Both coefficients are insignificant.

	(1) Premium	(2) Premium	(3) Premium
Reverse repo facility rate	-0.0101	-2.2225***	-2.1755***
	(0.070)	(0.446)	(0.372)
Reverse repo facility trade amounts		0.1498***	0.1186***
		(0.048)	(0.047)
lnVIX		0.8326***	0.7957***
		(0.321)	(0.287)
10 year - 2 year yield spread		0.2183	0.0898
		(0.161)	(0.115)
General collateral financing repo rate		1.6331***	1.4591***
		(0.406)	(0.358)
Constant	2.7820***	-0.1528	1.1022
	(0.093)	(1.051)	(0.886)
Crisis and maturity fixed effects	No	No	Yes
No. Observations:	14952	13076	13076
R-squared:	0.000	0.051	0.061
Adj. R-squared:	0.000	0.051	0.060

I use Newey-West standard errors with 6 lags. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level. The sample period is 23 September 2013-2021 Q2 and the frequency is daily. The data source for the premium (in basis points) is the same as in figure 3.2a. The reverse repo facility rate and trade amount data are provided by the New York FED and can be downloaded here: https://www.newyorkfed.org/markets/omo_transaction_data#rrp. I take daily averages. The rate is in percent, the trade amount in billion US dollars. I deflate the trade amounts using the "Consumer Price Index for All Urban Consumers: All Items in U.S. City Average", which can be downloaded from FRED. I set the index to 1 when the trade amount series starts on the 23 September 2013. The VIX data are taken from FRED. The 10 year and 2 year yields (par yields of seasoned bonds in percent) are taken from the FED yield curve, which is an updated version of the original Gürkaynak-Sack-Wright curve (Gürkaynak et al. (2010)), https://www.federalreserve.gov/data/tips-yield-curve-and-inflation-compensation.htm. The general collateral financing repo rate (in percent) is provided by the Depository Trust and Clearing Corporation (DTCC) here: https://www.dtcc.com/charts/dtcc-gcf-repo-index. The crisis fixed effects capture the months October 2014, September 2019, and March 2020 and the maturity fixed effects each maturity of the on-the-run premia (1y, 2y, 3y, 5y, 7y, 10y, 20y, and 30y).

Table 3.4: Premium on reverse repo facility rate regression

issuance of the assets. Each period and the auction are characterised by a different price. The chart below illustrates the life cycle.

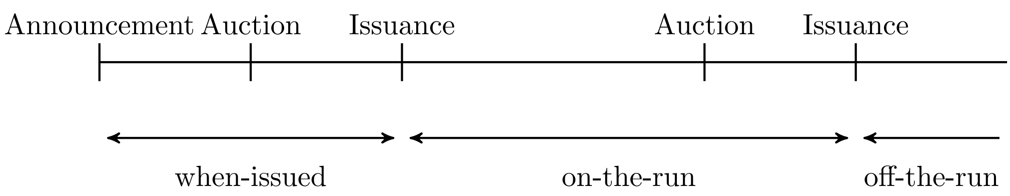


Figure 3.9: Life cycle

The when-issued market takes place after the auction of the security is announced but before it is issued. The when-issued market is important for price discovery (Durham and Perli (2022)). One possible interpretation of the prices q_t^2 and q_t^n is as the when-issued price of the respective assets. Thus, the model covers all three periods with its prices. The fourth important price during the life cycle, the auction price, is added to the model below. Since I know the on- and off-the-run prices, I can derive the auction price. I still look at the premium equilibrium.

I will first discuss the auction price of the one-period asset. A primary dealer taking part in the auction chooses to bid with a bundle consisting of price and quantity which I denote by $\left\{p_t^{1,A}, A_t^{1,A}\right\}$. The bundle specifies the quantity the primary dealer wants to buy and the price he is willing to pay. For each possible price $p_t^{1,A}$ that could be chosen, the optimal quantity is given by the solution to the following maximisation problem:

$$\begin{aligned}
&\max_{A_t^{1,A}} - p_t^{1,A} A_t^{1,A} + (1-\sigma)\left[p_t^n A_t^n - \kappa(A_t^n) + \delta(A_t^{1,A} - A_t^n)\right] + \sigma\delta A_t^{1,A} \\
&\quad \text{s.t.} \\
&\quad A_t^{1,A} \geq A_t^n.
\end{aligned}$$

If the primary dealer buys an asset at the auction, he must pay the price $p_t^{1,A}$ today. With probability $(1-\sigma)$ he can enter the market in the same subperiod and sell the quantity A_t^n of the asset. With probability σ he will hold the asset until maturity and consume the coupon. I know that in the premium equilibrium the constraint is not binding. Therefore, $p_t^{1,A} = \delta$. The price must equal the marginal utility of the asset. Otherwise there is infinite or no demand. Bidding will therefore drive the price to this value. The supply side is given by the Treasury, which auctions the amount $I_t^n = \mathcal{I}$. Therefore, in equilibrium, $A_t^{1,A} = I_t^n = \mathcal{I}$ and $p_t^{1,A} = \delta$.

The maximisation problem for the two-period asset is analogous:

$$\begin{aligned}
\max_{p_t^{2,A}, A_t^{2,A}} & - p_t^{2,A} A_t^{2,A} + (1-\sigma)\Big\{p_t^2 A_t^2 - \kappa(A_t^2) + \delta(A_t^{2,A} - A_t^2) \\
& + \beta(1-\sigma)\left[p_{t+1}^f A_{t+1}^{f,l} - \kappa(A_{t+1}^{f,l}) + \delta(A_t^{2,A} - A_t^2 - A_{t+1}^{f,l})\right] + \beta\sigma\delta(A_t^{2,A} - A_t^2)\Big\} \\
& + \sigma\left\{\delta A_t^{2,A} + \beta(1-\sigma)\left[p_{t+1}^f A_{t+1}^{f,h} - \kappa(A_{t+1}^{f,h}) + \delta(A_t^{2,A} - A_{t+1}^{f,h})\right] + \beta\sigma\delta A_t^{2,A}\right\} \\
& \text{s.t.} \\
& A_t^{2,A} \geq A_t^2 \\
& A_t^{2,A} \geq A_{t+1}^{f,h} \\
& A_t^{2,A} - A_t^2 \geq A_{t+1}^{f,l}.
\end{aligned}$$

I know that in the premium equilibrium, the last constraint is binding. The others are not.

The Treasury auctions the stock $I_t^2 = \mathcal{I}$. Therefore in equilibrium $A_t^{2,A} = I_t^2 = \mathcal{I}$ and $p_t^{2,A} = \delta + (1-\sigma)\beta\left\{(1-\sigma)[p_{t+1}^f - \kappa'(A^{f,l})] + \sigma\delta\right\} + \sigma\beta\delta$.

After adding the auction price, I can now summarise and compare all the prices. In the case of the one-period asset, these are

$$
\begin{aligned}
p_t^{1,A} &= \delta \\
q_t^n &= (\delta + g) \\
p_t^n &= \beta(1 + r_t)(\delta + g).
\end{aligned}
$$

In the case of the two-period asset, these are

$$
\begin{aligned}
p_t^{2,A} &= \delta + (1-\sigma)\beta\left\{(1-\sigma)[p_{t+1}^f - \kappa'(A^{f,l})] + \sigma\delta\right\} + \sigma\beta\delta \\
q_t^2 &= (1+\beta)(\delta + g) \\
p_t^2 &= \beta(1 + r_t)(1+\beta)(\delta + g) \\
p_t^f &= \frac{\sigma}{1 - (1-\sigma)\beta(1+r_t)}\beta(1 + r_t)(\delta + g).
\end{aligned}
$$

In addition to the empirical regularity that the on-the-run price is above the off-the-run price, the model explains two other regularities about prices as pointed out in the beginning of this section. First, a detailed analysis by Durham and Perli (2022) showed that when-issued prices carry a premium compared to already issued Treasuries. In their comparison, they include both on- and off-the-run Treasuries. Second, the secondary market price is known to be higher than the auction price (Goldreich (2007), Spindt and Stolz (1992), and Fleming et al. (2022)). See appendix section 3.12.3 for an illustration. The following proposition shows that these two stylised facts also hold in the premium equilibrium.

Proposition 5. *In the premium equilibrium*

$$
\begin{aligned}
q_t^2 &> p_t^2 \\
q_t^n &> p_t^n \\
q_t^n &> p_t^f.
\end{aligned}
$$

and

$$
\begin{aligned}
p_t^2 &> p_t^{2,A} \\
p_t^n &> p_t^{1,A}.
\end{aligned}
$$

Proof. See appendix 3.12.2.4. □

I will explain the first of the two stylised facts first. It follows from the equations that the spot price (in the case of no fail) is the discounted when-issued price. This can be explained from the seller's perspective. When selling a when-issued contract, the seller receives the when-issued price tomorrow and can deposit it. This income has to cover the spot price he has to pay tomorrow, which in equilibrium is equal to the collateral value he has to build up today. Therefore the spot price in case of no fail is the discounted when-issued price. Compared to off-the-run prices, the when-issued price is even higher, because the off-the-run price takes into account the probability of a settlement fail.

To explain the second stylised fact, the most natural comparison is to compare the auction to the on-the-run price in our model. The auction price is given by the marginal utility of the last unit purchased. The last unit remains in the inventory of the primary dealer and he consumes the coupon δ, except it is sold as an off-the-run asset of a constrained primary dealer. The empirical results of Fleming et al. (2022) suggest that some of the Treasuries remain in the portfolios of the primary dealers until maturity, while the other part is sold. This is consistent with our observation and is reflected in the auction price.

On the other hand, the on-the-run price is determined by the value of the asset to participants in the secondary market. In our case, the buyer is the ultimate owner of the asset, so his marginal utility determines the price. In equilibrium, the spot price is above the coupon rate δ because the buyer has a higher marginal valuation of the asset. The on-the-run prices are therefore higher than the auction prices.

I summarise the above discussion as follows: The on-the-run prices are driven by the valuation of the buyer of the asset in the secondary market. The auction prices also contains the valuation of the asset if it stays in the inventory of the primary dealer until maturity. The latter valuation is lower than the former. Therefore the primary market prices lie below the secondary market prices. The when issued prices are higher than the spot prices due to the cost of collateralisation.

3.11 Conclusion

I developed a model of the on-the-run phenomenon. I provide a novel explanation using inventory which I also test empirically. The model was then used to discuss broad access to central bank facilities. The analysis is motivated by the current discussion on how to reform the market (see e.g. Duffie (2023)) and the increase in intermediation and participation by non-bank financial institutions (Eren and Wooldridge (2021)). The latter raises the question of why only a limited number of participants currently have access to central bank facilities.[51] I analysed the implications of providing broad access to a reverse repo or deposit facility. Finally, I added to the literature by extending the model to cover the full life cycle of Treasuries. It includes the auction price, the when-issued price, and the on- and off-the-run prices (all four relevant prices), and can accommodate stylised facts about the relationship between them.

To explain the premium, I endogenised settlement fails and found, consistent with the data, that they are more frequent for off-the-run Treasuries. The probability of a settlement fail is higher for off-the-run Treasuries because they are longer in the market and more dispersed. This leads to a preference for and premium on on-the-run Treasuries. This also explains why the premium is always on the on-the-run Treasury, an aspect that Vayanos and Weill (2008) could not explain. I also test my theory in the data. I show that the premium is higher in times when the probability of failure of off-the-run Treasuries is higher. The probability of fails with off-the-run Treasuries is higher, when the inventories of primary dealers are lower.

Summarising, the on-the-run premium is a symptom that the Treasury market is not frictionless, and inventory risk is higher for off-the-run Treasuries, making them unattractive. Future research should analyse whether the fundamental dynamics described imply a self-fulfilling dynamic that adds to the size of the premium.

The second part focused on the impact of the facility rate when broad access is provided. The result is that an increase in the facility rate leads to a reduction in the cost of trading. The facility complements a trade in the sense that it provides a certain return on liquid funds between trades for those who have access to it. I show that access stimulates trading and prices rise. An increase in the facility rate decreases the premium, which I also confirm in the data. Interestingly, if the facility rate increases, fails increase as more off-the-run Treasuries are traded in contracts promising their delivery but fewer are available. Also, giving access to all institutions is not an act which equalises benefits. Only primary dealers benefit from an increase in the facility rate, not those who gained access.

An interesting way to extend the model would be to take into account more specific characteristics of non-bank financial institutions. This could be, for example, leverage in the case of a hedge fund. In addition, given the current discussions on how to restructure the Treasury market (see e.g. Duffie et al. (2021)), it would be interesting to explore the effects of broad access to a repo facility to complement my analysis.

[51] See for example the FAQ on the repo and the reverse repo facility of the New York FED for eligibility criteria, https://www.newyorkfed.org/markets/repo-agreement-ops-faq and https://www.newyorkfed.org/markets/rrp_faq.

3.12 Appendix

3.12.1 Other equilibria

Until now I focused on the case where $\lambda_t^2 = \lambda_t^n = \lambda_t^{f,h} = 0$. This means that primary dealers are unconstrained if they still have the full inventory of assets i. I showed that I can always find issuance sizes $\mathcal{I}$ where such an equilibrium exists. For the sake of completeness I can also think of equilibria where λ_t^2, λ_t^n, and $\lambda_t^{f,h}$ are non-zero. In this case, the quantity of the corresponding assets sold is $\mathcal{I}$. The prices in each of these equilibria are the same as in the premium equilibrium 3.6 as long as $\lambda_t^{f,l} > 0$, which I need for a positive premium. Since the dynamics and intuition are exactly the same in these equilibria as well, I don't discuss these equilibria in the following, as no additional insights are gained. The only results that would change are the effects of central bank access on quantities. In these other equilibria, there are no quantity effects on assets for which constraints are binding, and only the Lagrange multipliers would change in magnitude.

3.12.2 Proofs

3.12.2.1 Proof of proposition 1

Proof. Suppose it is not the case that $\mathcal{I} < A_{t-1}^2 + A_t^{f,h}$, or $\lambda_t^{f,l} > 0$. Then $\lambda_t^{f,l} = \widetilde{\lambda}_t^{f,l} = 0$. From the first-order conditions of the seller and the buyer and $\omega_t^f = p_t^f$, it follows that $p_t^n = p_t^f = \beta^2(1 + r_t)(\delta + g)$. Therefore, there is no premium, $\Delta_t = 0$. The case $\lambda_t^{f,l} = 0$ and $\widetilde{\lambda}_t^{f,l} > 0$ is not possible. If $\lambda_t^{f,l} = 0$, then it must be that $\widetilde{\lambda}_t^{f,l} = \widetilde{\lambda}_t^{f,h} = 0$ given the optimal behavior of the seller which takes into account the profitable amount to be sold by the primary dealer. Also the case where $\lambda_{t+1}^{f,l} > 0$ and $\widetilde{\lambda}_{t+1}^{f,l} > 0$ is not possible. If $\lambda_{t+1}^{f,h} = 0$ and $\lambda_{t+1}^{f,l} > 0$, then $a_t^f > A_t^{f,l}$, and it follows that $\widetilde{\lambda}_{t+1}^{f,l} = 0$. $\square$

3.12.2.2 Proof of proposition 2

Proof. First, I show that for an equilibrium with trade in all assets to exist, I need $\sigma > \frac{1-\beta(1+r_t)}{\beta(1+r_t)} \frac{\delta}{g}$. In equilibrium $p_t^f = \kappa'(A_t^{f,h}) + \delta = \frac{\sigma}{1-(1-\sigma)\beta(1+r_t)}\beta(1 + r_t)(\delta + g)$. For $\kappa'(A_t^{f,h}) > 0$ and $A_t^{f,h} > 0$ I must have that $\frac{\sigma}{1-(1-\sigma)\beta(1+r_t)}\beta(1 + r_t)(\delta + g) - \delta > 0$, or $\sigma > \frac{1-\beta(1+r_t)}{\beta(1+r_t)} \frac{\delta}{g}$.

Second, I show that $A_t^2 > A_t^n > A_t^{f,h}$. The first order conditions of the primary dealer with respect to the assets maturing in one period are $p_t^n = \kappa'(A_t^n) + \delta$ and $p_t^f = \kappa'(A_t^f) + \delta$. As $p_t^n > p_t^f$ and $\kappa(A_t^i)$ being a strictly convex function it follows that $A_t^n > A_t^f$. In addition, I can show that $A_t^2 > A_t^n$. The first order condition of the primary dealer with respect to the new two-period asset is:

$$\begin{aligned}
\kappa'(A_t^2) - \beta(1-\sigma)\kappa'(\mathcal{I} - A_t^2) &= p_t^2 - (1+\beta)\delta - \beta(1-\sigma)\kappa'(A_{t+1}^{f,h}) \\
\kappa'(A_t^2) - \beta(1-\sigma)\kappa'(\mathcal{I} - A_t^2) &= (1+\beta)p_t^n - (1+\beta)\delta - \beta(1-\sigma)\kappa'(A_{t+1}^{f,h}) \\
\kappa'(A_t^2) - \beta(1-\sigma)\kappa'(\mathcal{I} - A_t^2) &= (1+\beta)\kappa'(A_t^n) - \beta(1-\sigma)\kappa'(A_{t+1}^{f,h}).
\end{aligned}$$

Further rearrangement yields $\kappa'(A_t^2) - \kappa'(A_t^n) = \beta\kappa'(\mathcal{I} - A_t^2) + \beta[\kappa'(A_t^n) - \kappa'(A_{t+1}^{f,h})] + \beta\sigma[\kappa'(A_{t+1}^{f,h}) - \kappa'(\mathcal{I} - A_t^2)] > 0$. Therefore $A_t^2 > A_t^n$.

Lastly I show that I can always find an $\mathcal{I}$, where the equilibrium exists. In equilibrium $(A_{t-1}^2 + A_t^{f,h}) > \mathcal{I} > max(A_t^2, A_t^n, A_t^{f,h})$. I know that $A_t^2 > A_t^n > A_t^{f,h}$. Therefore it follows that in equilibrium $(A_{t-1}^2 + A_t^{f,h}) > \mathcal{I} > A_t^2$. From the equilibrium conditions it follows that

$$\begin{aligned}
A_t^n &= \kappa'^{-1}\left[\beta(1+r_t)(\delta+g) - \delta\right] \\
A_t^{f,h} &= \kappa'^{-1}\left[\frac{\sigma}{1-(1-\sigma)\beta(1+r_t)}\beta(1+r_t)(\delta+g) - \delta\right] \\
\kappa'(A_{t-1}^2) &= W + \beta(1-\sigma)\kappa'(\mathcal{I} - A_{t-1}^2)
\end{aligned}$$

where $W \equiv \left[(1+\beta) - \beta(1-\sigma)\frac{\sigma}{1-(1-\sigma)\beta(1+r_t)}\right]\beta(1+r_t)(\delta+g) - (1+\beta\sigma)\delta$.

The left-hand side of the third equation, $f_1(x) = \kappa'(x)$, is a strictly increasing function with $x \in [0, \mathcal{I}]$ and $min(f_1(x)) = 0$ and $max(f_1(x)) = \kappa'(\mathcal{I})$. The right-hand side of the third equation, $f_2(x) = W + \beta(1-\sigma)\kappa'(\mathcal{I} - x)$, is a strictly decreasing function with $x \in [0, \mathcal{I}]$ and $min(f_2(x)) = W$ and $max(f_2(x)) = W + \beta(1-\sigma)\kappa'(\mathcal{I})$. Therefore for the equilibrium to exist it must be that $\kappa'(\mathcal{I}) > W$, or $\mathcal{I} > \kappa'^{-1}(W)$. By continuity it then follows that $\mathcal{I} > A_{t-1}^2$. As $\mathcal{I} - A_{t-1}^2(\mathcal{I})$ is a continuous function (as $\kappa'(A_t^i)$ and $\kappa'^{-1}(A_t^i)$ are both continuous) with minimal value 0 for $\mathcal{I} = \kappa'^{-1}(W)$ I can always find an $\mathcal{I}$ where $\mathcal{I} - A_{t-1}^2(\mathcal{I}) < A_t^{f,h}$ for any $A_t^{f,h} > 0$. □

3.12.2.3 Proof of proposition 4

Proof. First, I know that in equilibrium $f_t^n = 0$ and $f_t^f = \frac{(1-\sigma)p_t^f(a_t^f - I_t^f)}{p_t^f a_t^f} = (1-\sigma)\left(1 - \frac{\mathcal{I} - a_t^2}{a_t^f}\right)$. Therefore, trivially, $f_t^f > 0 = f_t^n$. Second, from comparing $p_t^n = \beta(1+r_t)(\delta+g)$ and $p_t^f = \frac{\sigma}{1-(1-\sigma)\beta(1+r_t)}\beta(1+r_t)\beta(\delta+g)$ it follows immediately that $p_t^n > p_t^f$. Third, $p_t^n > p_t^f$ also implies that $A_t^n > A_t^{f,h}$ as $A_t^n = \kappa'^{-1}(p_t^n - \delta)$ and $A_t^{f,h} = \kappa'^{-1}(p_t^f - \delta)$. Lastly, as $\lambda_t^{f,h} = 0$ and $\lambda_t^{f,l} > 0$ it follows that $A_t^{f,h} > A_t^{f,l}$. □

3.12.2.4 Proof of proposition 5

Proof. First, as $\beta(1+r_t) > 0$, it follows that $q_t^2 > p_t^2$ and $q_t^n > p_t^n$. Also, from proposition 4, I know that $p_t^n > p_t^f$ and therefore $q_t^n > p_t^f$. Second, from the first order condition of the primary dealer (see section 3.4.1) I know that $p_t^2 = \kappa'(A_t^2) + \delta + \beta\delta + \beta(1-\sigma)\lambda_{t+1}^f$ where $\lambda_{t+1}^f = p_{t+1}^f - \kappa'(A^{f,l}) - \delta$. It follows immediately that $p_t^2 > p_t^{2,A}$. Also from the first order condition of the primary dealer, I know that $p_t^n = \kappa'(A_t^n) + \delta$. It follows immediately that $p_t^n > p_t^{1,A}$. □

3.12.2.5 Proof of corollary 1

Proof. I assume an unanticipated permanent increase in the interest rate that occurs at the beginning of the contract market in $t = \widetilde{t} - 1$. The stock of available off-the-run assets in the next spot market is given, and it can reach a new steady state only after a period.

First, I show that an increase in the facility rate r_t raises all prices. I know that in equilibrium, prices are given by $p_t^n = \beta(1+r_t)(\delta+g)$, $p_t^2 = \beta(1+r_t)(1+\beta)(\delta+g)$, $p_t^f = \frac{\sigma}{1-(1-\sigma)\beta(1+r_t)}\beta(1+r_t)(\delta+g)$. It follows that

$$\begin{aligned}
\frac{dp_t^n}{dr_{\widetilde{t}-1}} &= \beta(\delta+g) > 0 \\
\frac{dp_t^2}{dr_{\widetilde{t}-1}} &= \beta(1+\beta)(\delta+g) > 0 \\
\frac{dp_t^f}{dr_{\widetilde{t}-1}} &= \left[\frac{\sigma}{1-(1-\sigma)\beta(1+r_t)} + \frac{\sigma(1-\sigma)\beta(1+r_t)}{[1-(1-\sigma)\beta(1+r_t)]^2}\right]\beta(\delta+g) > 0
\end{aligned}$$

for all $t \geq \widetilde{t}$.

Second, I show that the on-the-run premium decreases given an increase in r_t. In equilibrium $\frac{dp_t^n}{dr_{\widetilde{t}-1}} < \frac{dp_t^f}{dr_{\widetilde{t}-1}}$ if $\frac{\sigma}{1-(1-\sigma)\beta(1+r_t)} + \frac{\sigma(1-\sigma)\beta(1+r_t)}{[1-(1-\sigma)\beta(1+r_t)]^2} > 1$ which is the case if $\sigma > \frac{[1-\beta(1+r_t)]^2}{[\beta(1+r_t)]^2} \approx 0$.[52] The premium is given by $\Delta_t = p_t^n - p_t^f$. Therefore $\frac{d\Delta_t}{dr_{\widetilde{t}-1}} = \frac{dp_t^n}{dr_{\widetilde{t}-1}} - \frac{dp_t^f}{dr_{\widetilde{t}-1}} < 0\, \forall t \geq \widetilde{t}$.

Third, I discuss the impact on the quantities. I know that in equilibrium quantities are determined by $\kappa'(A_t^n) = p_t^n - \delta$, $\kappa'(A_t^2) - \beta(1-\sigma)\kappa'(I_t^2 - A_t^2) = p_t^2 - \beta(1-\sigma)p_{t+1}^f - (1+\beta\sigma)\delta$,

[52] Let me define $k \equiv (1-\sigma)\beta(1+r_t)$. It follows: $\frac{\sigma}{1-k} + \frac{\sigma k}{(1-k)^2} > 1$, or $\sigma > (1-k)^2$. Further rearranging yields $0 > 1 - 2\beta(1+r_t) + (1-\sigma)[\beta(1+r_t)]^2$, or $\sigma > \frac{[1-\beta(1+r_t)]^2}{[\beta(1+r_t)]^2} \approx 0$.

and $\kappa'(A_t^{f,h}) = p_t^f - \delta$. It follows that

$$\frac{d\kappa'(A_t^n)}{dr_{\tilde{t}-1}} = \frac{dp_t^n}{dr_{\tilde{t}-1}} > 0$$
$$\frac{d\kappa'(A_t^{f,h})}{dr_{\tilde{t}-1}} = \frac{dp_t^f}{dr_{\tilde{t}-1}} > 0$$

for all $t \geq \tilde{t}$. Therefore $\frac{dA_t^n}{dr_{\tilde{t}-1}} > 0$ and $\frac{dA_t^{f,h}}{dr_{\tilde{t}-1}} > 0$ for all $t \geq \tilde{t}$. Also

$$\frac{d\kappa'(A_t^2)}{dr_{\tilde{t}-1}} - \beta(1-\sigma)\frac{d\kappa'(I_t^2 - A_t^2)}{dr_{\tilde{t}-1}} = \frac{dp_t^2}{dr_{\tilde{t}-1}} - \beta(1-\sigma)\frac{dp_{t+1}^f}{dr_{\tilde{t}-1}}.$$

It follows that $\frac{dA_t^2}{dr_{\tilde{t}-1}} > 0$ for all $t \geq \tilde{t}$ iff $\frac{dp_t^2}{dr_{\tilde{t}-1}} > \beta(1-\sigma)\frac{dp_{t+1}^f}{dr_{\tilde{t}-1}}$. $\frac{dp_t^2}{dr_{\tilde{t}-1}} > \beta(1-\sigma)\frac{dp_{t+1}^f}{dr_{\tilde{t}-1}}$ if $\sigma > \tilde{\sigma}$.

The value of $\tilde{\sigma}$ is derived as follows:

$\frac{dp_t^2}{dr_{\tilde{t}-1}} > \beta(1-\sigma)\frac{dp_{t+1}^f}{dr_{\tilde{t}-1}}$ iff $(1+\beta) > \beta(1-\sigma)\left[\frac{\sigma}{1-(1-\sigma)\beta(1+r_t)} + \frac{\sigma(1-\sigma)\beta(1+r_t)}{[1-(1-\sigma)\beta(1+r_t)]^2}\right]$, or $(1-\beta(1+r_t))^2 > \frac{\sigma}{(1-\sigma)}\left[(1-\sigma)\left(\frac{\beta}{(1+\beta)} + (\beta(1+r_t))^2\right) - 1\right]$, where $(1-\beta(1+r_t))^2 \approx 0$. There are two values of σ which solve the equation $(1-\beta(1+r_t))^2 = \frac{\sigma}{(1-\sigma)}\left[(1-\sigma)\left(\frac{\beta}{(1+\beta)} + (\beta(1+r_t))^2\right) - 1\right]$. Below the lower root and above the upper root the right-hand side is lower than the left-hand side. I define $\tilde{\sigma}$ as the larger root. Note that using empirically reasonable values, it can only for a limited range of small values of σ be that $\frac{dp_t^2}{dr_{\tilde{t}-1}} > \frac{dp_{t+1}^f}{dr_{\tilde{t}-1}}$ is not true. For example if $\beta = 0.96$ and $r_{\tilde{t}-1} = 0.01$, then only for $\sigma \in (0.003, 0.25)$ it follows that $\frac{dp_t^2}{dr_{\tilde{t}-1}} < \frac{dp_{t+1}^f}{dr_{\tilde{t}-1}}$.

Next I look at the quantity of off-the-run assets traded by constrained dealers, which is given by $A_t^{f,l} = \mathcal{I} - A_{t-1}^2$. It follows that $\frac{dA_t^{f,l}}{dr_{\tilde{t}-1}} = 0$ for $t = \tilde{t}$ because $\frac{dA_{\tilde{t}-1}^2}{dr_{\tilde{t}-1}} = 0$. For all $t > \tilde{t}$ it follows that $\frac{dA_t^{f,l}}{dr_{\tilde{t}-1}} < 0$ if $\frac{dA_t^2}{dr_{\tilde{t}-1}} > 0$ which is true if $\sigma > \tilde{\sigma}$. From the above results, it directly follows that $\frac{d(A_t^2 + A_t^n + \sigma A_t^{f,h} + (1-\sigma)A_t^{f,l})}{dr_{\tilde{t}-1}} > 0$ for $t = \tilde{t}$ and $\forall t > \tilde{t}$ if $\sigma > \tilde{\sigma}$. Decreases in the trading of off-the-run assets by constrained primary dealers are due to equivalent increases in the trading of the asset in the on-the-run period.

Lastly, the off-the-run settlement failure rate is given by $f_t^f = \frac{(1-\sigma)p_t^f(a_t^f - I_t^f)}{p_t^f a_t^f} = (1-\sigma)\left(1 - \frac{\mathcal{I} - a_{t-1}^2}{a_t^f}\right)$. It follows that $\frac{df_t^f}{dr_{\tilde{t}-1}} > 0$ for $t = \tilde{t}$ because $\frac{dp_t^f}{dr_{\tilde{t}-1}} > 0$ for $t \geq \tilde{t}$ and $\frac{\partial A_{\tilde{t}-1}^2}{\partial r_{\tilde{t}-1}} = \frac{\partial a_{\tilde{t}-1}^2}{\partial r_{\tilde{t}-1}} = 0$. For $\forall t > \tilde{t}$ it follows that $\frac{df_t^f}{dr_{\tilde{t}-1}} > 0$ iff $\frac{\partial A_{t-1}^2}{\partial r_{\tilde{t}-1}} = \frac{\partial a_{t-1}^2}{\partial r_{\tilde{t}-1}} > 0$ which is the case if $\sigma > \tilde{\sigma}$. $\square$

3.12.2.6 Proof of corollary 2

Proof. I look at the impact of an increase in the interest rate r_t (see details below) on the lifetime values of the buyer, the seller, and the primary dealer. There are two groups of primary dealers: One group of primary dealers has an inventory of $\mathcal{I}$ off-the-run assets and another has one of $\mathcal{I} - A_t^2$. I take the lifetime value of both groups and average according to their share in the population. To ease notation, let me define $V_{\tilde{t}-1}^{D,a} \equiv (1-\sigma)V_{\tilde{t}-1}^{D}(\mathcal{I} - A_{\tilde{t}}^2) + \sigma V_{\tilde{t}-1}^{D}(\mathcal{I})$. The lifetime values are given by

$$
\begin{aligned}
V_{\tilde{t}-1}^{b} &= \sum_i \left[\beta(\delta+g) - \beta q_{\tilde{t}-1}^{i}\right] a_{\tilde{t}}^{i} + \beta^2(\delta+g)a_{\tilde{t}}^{2} - \beta(1-\sigma)[(\delta+g) - \omega_{\tilde{t}-1}^{f}](a_{\tilde{t}}^{f} - I_{\tilde{t}}^{f}) + \beta V_{\tilde{t}}^{b} \\
V_{\tilde{t}-1}^{s} &= \sum_i \left\{-\omega_{\tilde{t}-1}^{i} + \beta\left[q_{\tilde{t}-1}^{i} - p_{\tilde{t}}^{i} + \omega_{\tilde{t}-1}^{i}\right](1+r_t)a_{\tilde{t}}^{i}\right\} - \beta(1-\sigma)[\omega_{\tilde{t}-1}^{f} - p_{\tilde{t}}^{f}](1+r_t)(a_{\tilde{t}}^{f} - I_{\tilde{t}}^{n}) \\
&\quad + \beta V_{\tilde{t}-1}^{s} \\
V_{\tilde{t}-1}^{D,a} &= \beta\{(1-\sigma)^2[p_{\tilde{t}}^{f}(I_{\tilde{t}-1}^{2} - A_{\tilde{t}-1}^{2}) - \kappa(I_{\tilde{t}-1}^{2} - A_{\tilde{t}-1}^{2}) - \delta(I_{\tilde{t}-1}^{2} - A_{\tilde{t}-1}^{2})] \\
&\quad + (1-\sigma)\sigma[p_{\tilde{t}}^{f}A_{\tilde{t}}^{f,h} - \kappa(A_{\tilde{t}}^{f,h}) - \delta A_{\tilde{t}}^{f,h}] + (1-\sigma)\sum_{j\in\{n,2\}}\left[p_{\tilde{t}}^{j}A_{\tilde{t}}^{j} - \kappa(A_{\tilde{t}}^{j}) - \delta A_{\tilde{t}}^{j}\right] \\
&\quad + \delta(I_{\tilde{t}}^{n} + I_{\tilde{t}}^{2} + I_{\tilde{t}-1}^{2} - (1-\sigma)A_{\tilde{t}-1}^{2})\} + \beta V_{\tilde{t}}^{D,a}.
\end{aligned}
$$

I assume that there is an unanticipated permanent increase in the interest rate at the beginning of the contract market in $t = \tilde{t} - 1$. The stock of available off-the-run assets in the next spot market is given and can reach a new steady state only after a period. New steady state values have no time index. I can simplify the buyer's and seller's profit using the first-order conditions in equilibrium. This yields

$$
\begin{aligned}
V_{\tilde{t}-1}^{b} &= \beta(1-\sigma)[(\delta+g) - p_{\tilde{t}}^{f}]I_{\tilde{t}}^{f} + \frac{\beta^2}{1-\beta}(1-\sigma)\left[(\delta+g) - p^f\right]I^f \\
V_{\tilde{t}-1}^{s} &= 0.
\end{aligned}
$$

The derivatives are

$$
\begin{aligned}
\frac{dV_{\tilde{t}-1}^{b}}{dr_{\tilde{t}-1}} &= -\beta(1-\sigma)\frac{dp_{\tilde{t}}^{f}}{dr_{\tilde{t}-1}}(\mathcal{I} - a_{\tilde{t}-1}^{2}) - \frac{\beta^2}{1-\beta}(1-\sigma)\left\{\frac{dp^f}{dr_{\tilde{t}-1}}(\mathcal{I} - a^2) + \left[(\delta+g) - p^f\right]\frac{da^2}{dr_{\tilde{t}-1}}\right\} \\
\frac{dV_{\tilde{t}-1}^{d}}{dr_{\tilde{t}-1}} &= 0
\end{aligned}
$$

$$\frac{dV^{D,a}_{\widetilde{t}-1}}{dr_{\widetilde{t}-1}} = \beta\Bigg\{(1-\sigma)^2\frac{dp^f_{\widetilde{t}}}{dr_{\widetilde{t}-1}}(\mathcal{I}-A^2_{\widetilde{t}-1}) + (1-\sigma)\sigma\frac{d[\kappa'(A^{f,h}_{\widetilde{t}})A^{f,h}_{\widetilde{t}} - \kappa(A^{f,h}_{\widetilde{t}})]}{dr_{\widetilde{t}-1}}$$

$$+ (1-\sigma)\sum_{j\in\{n,2\}}\frac{d[\kappa'(A^j_{\widetilde{t}})A^j_{\widetilde{t}} - \kappa(A^j_{\widetilde{t}})]}{dr_{\widetilde{t}-1}}\Bigg\}$$

$$+ \frac{\beta^2}{1-\beta}\Bigg\{(1-\sigma)^2\frac{d[\kappa'(\mathcal{I}-A^2)(\mathcal{I}-A^2) - \kappa(\mathcal{I}-A^2) - \beta(1-\sigma)\kappa'(\mathcal{I}-A^2_{\widetilde{t}-1})\mathcal{I}]}{dr_{\widetilde{t}-1}}$$

$$+ (1-\sigma)\sigma\frac{d[\kappa'(A^{f,h})A^{f,h} - \kappa(A^{f,h})]}{dr_{\widetilde{t}-1}} + (1-\sigma)\sum_{j\in\{n,2\}}\frac{d[\kappa'(A^j)A^j - \kappa(A^j)]}{dr_{\widetilde{t}-1}}\Bigg\}.$$

I can simplify the last equation. This yields

$$\frac{dV^{D,a}_{\widetilde{t}-1}}{dr_{\widetilde{t}-1}} = \beta(1-\sigma)^2\frac{dp^f_{\widetilde{t}}}{dr_{\widetilde{t}-1}}(\mathcal{I}-A^2_{\widetilde{t}-1}) + \frac{\beta}{1-\beta}\Bigg\{\beta(1-\sigma)^2\kappa''(\mathcal{I}-A^2)A^2\frac{\partial A^2}{\partial r_{\widetilde{t}-1}}$$

$$+ (1-\sigma)\sigma\kappa''(A^{f,h})A^{f,h}\frac{\partial A^{f,h}}{\partial r_{\widetilde{t}-1}} + (1-\sigma)\sum_{j\in\{n,2\}}\kappa''(A^j)A^j\frac{\partial A^j}{\partial r_{\widetilde{t}-1}}\Bigg\}.$$

Regarding the profits of the buyer, note that $(\delta+g)-p^f > 0$. If $\frac{\partial A^2}{\partial r_{\widetilde{t}-1}} = \frac{\partial a^2}{\partial r_{\widetilde{t}-1}} > 0$, which is the case if $\sigma > \widetilde{\sigma}$, then $\frac{dV^b_{\widetilde{t}-1}}{dr_{\widetilde{t}-1}} < 0$, $\frac{dV^s_{\widetilde{t}-1}}{dr_{\widetilde{t}-1}} = 0$, and $\frac{dV^{D,a}_{\widetilde{t}-1}}{dr_{\widetilde{t}-1}} > 0$. □

3.12.2.7 Proof of corollary 4

Proof. I analyse the impact if sellers without access gain access, meaning that their interest rate unanticipated permanently increases from r^m_t to r_t in the second subperiod in $t = \widetilde{t}-1$. From the equations in section 3.8.2 I immediately observe that the spot prices and quantities of all on-the-run assets traded by sellers who gain access increase, while the prices and quantities for the ones who already had access stay the same. All sellers now trade at the same prices $p^{j,a}_t$ and quantities $A^{j,a}_t$ for $j = 2$ and n. In the first period after the shock in $t = \widetilde{t}$, I immediately observe that this also holds for the off-the-run prices p^f_t as well as quantities if no inventory constraint binds, i.e. $A^{f,h}_t$. For this result, it is important to note that the probabilities of finding the assets have not yet changed because inventories are predetermined from the previous period. This also implies that the off-the-run quantities sold, if inventory constraints bind, $A^{f,l}_t$, stay the same in this first period. The on-the-run premium initially, i.e. in $t = \widetilde{t}$, decreases following the same reasoning as in corollary 2.

To analyse the off-the-run prices and quantities traded in the periods $t > \widetilde{t}$, I must take into account how the probabilities to find the assets change in the second period. Otherwise, the reasoning is analogous. I assume that settlement fails still occur, so that there is still a positive premium (as in section 3.8.1). The probability to find the assets is for all sellers σ. This is weakly lower than beforehand. If already before access $\mathbb{P}^a = \mathbb{P}^{na} = \sigma$, then the effects are the same as in the first period after the shock. Otherwise, the effects on the prices, the premium, and the quantities traded are ambiguous. $\square$

3.12.3 Figures

The following graphs show the stylised facts and irregularities that are discussed in the paper.

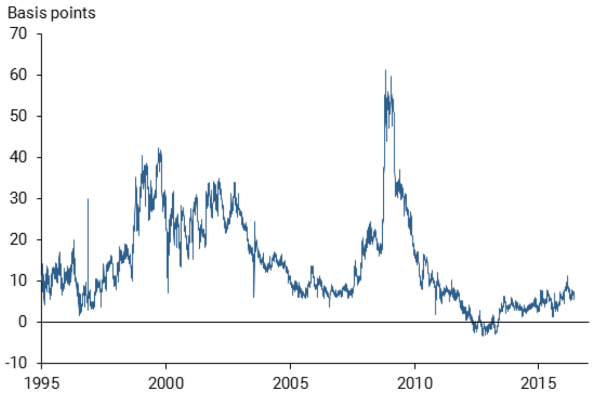


Figure 3.10: On-the-run premium on 10 year bonds

Figure 3.10 is taken from Christensen et al. (2017) and shows the on-the-run premium on 10 year Treasury bonds. The on-the-run yield is subtracted from the par yield of seasoned bonds.

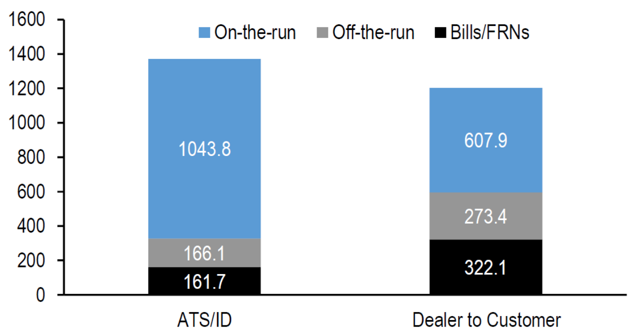


Figure 3.11: Treasury trading volumes by counterparty in billion USD

Figure 3.11 shows Treasury trading volumes by counterparty. ATS stands for automated trading system and ID for interdealer (U.S. Department of the Treasury (2022)). I can see that the trading volume in on-the-run Treasuries is much higher than in off-the-run Treasuries, regardless of the counterparty. This is also shown in the next chart.

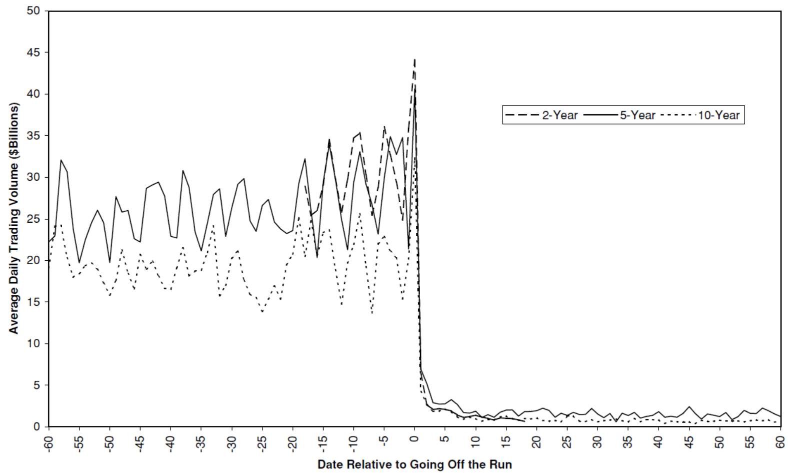


Figure 3.12: Trading volume for on-the-run and off-the-run Treasuries

Figure 3.12 shows the average trading volume for on- and off-the-run 2 year, 5 year, and 10 year Treasury notes relative to the date they went off-the-run (Barclay et al. (2006)).

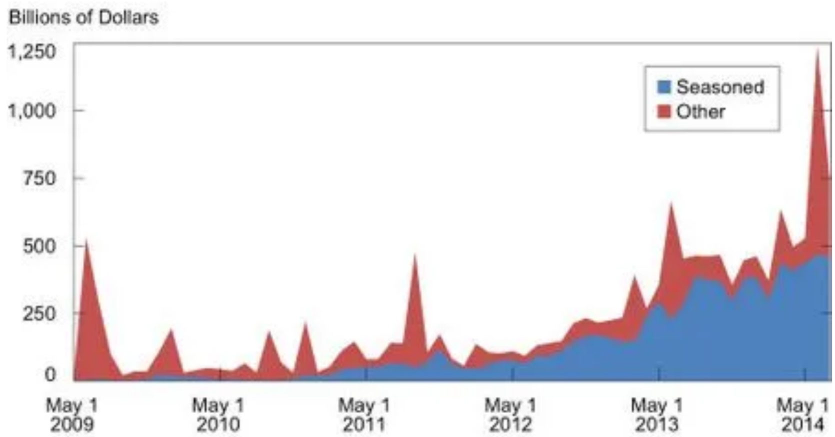


Figure 3.13: Fails in seasoned and other Treasuries

Figure 3.13, provided by Fleming et al. (2014), plots the cumulative gross fails by month in seasoned Treasuries and all other Treasuries (including on-the-run Treasuries). Seasoned Treasuries are Treasuries issued more than 180 days ago. Back in 2014, Fleming et al. (2014)

pointed out that there has been a steady increase in the number of fails of seasoned Treasuries over the past few years.

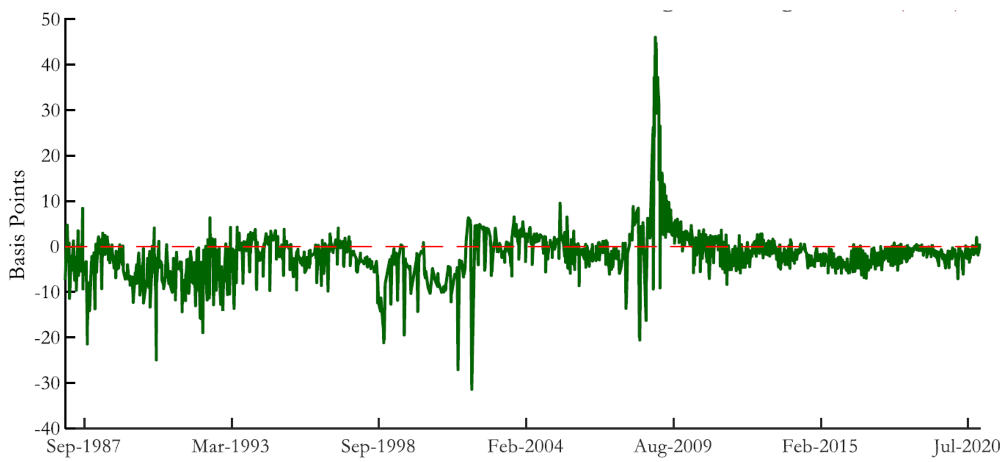


Figure 3.14: When-issued premium

The conditional when-issued premium, shown in figure 3.14, was measured by Durham and Perli (2022). The premium is computed by regressing the difference between the actual and the fitted yield on control variables and a dummy variable indicating whether the Treasury is when-issued.

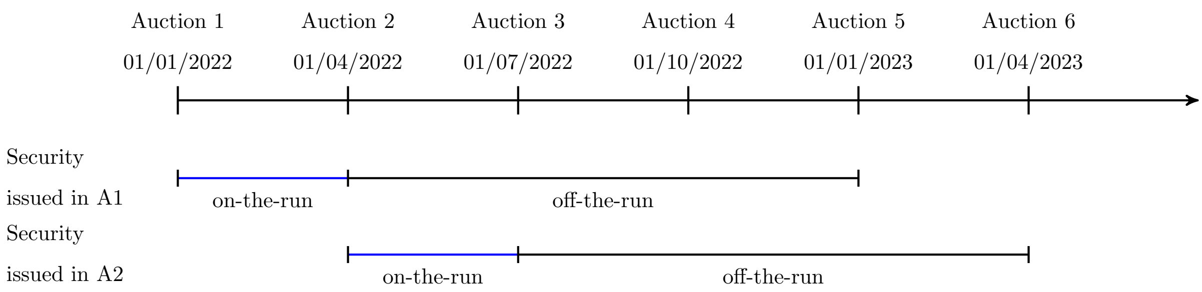
Time of Day	No. of Auctions	Underpricing (relative to when-issued transactions)	Standard Error	Underpricing (relative to when-issued bid quote)	Standard Error
12:50 PM–1:00 PM	178	0.32***	0.12	0.20**	0.12
12:30 PM–1:00 PM	178	0.40***	0.12	0.29***	0.12
1:00 PM–1:30 PM	176	0.32***	0.10	0.15*	0.10

Figure 3.15: Underpricing

Fleming et al. (2022) document in their paper that Treasury dealers appear to be compensated for taking inventory risks at the auction by price increases in subsequent weeks. This is evidence that auction prices are lower than secondary market prices after the auction. A direct comparison of primary and secondary market prices at the time of the auction has been done by Goldreich (2007) and Spindt and Stolz (1992), among others. Both show that the primary market price is lower than the secondary market price of the same security.[53] The table above from Goldreich (2007) compares the auction yields with the when-issued yields in the minutes before and after the auction. As Goldreich (2007) explains, for example, an underpricing of 0.32 basis points (first row, third column) is equivalent to 1.3 cents per 100$.

3.12.4 On-the-run cycle

The following figure illustrates the on-the-run cycle. An auction is held every quarter. The newly issued security is on-the-run until the next auction of assets with the same maturity. It then becomes off-the-run.



[53]Underpricing is also an observed phenomenon in Initial Public Offerings (IPO's). See for example Chambers and Dimson (2009).

Second Chapter

Transaction Costs, the Price of Convenience, and the Cross-Section of Safe Asset Returns

Transaction Costs, the Price of Convenience, and the Cross-Section of Safe Asset Returns[1]

co-authored with

Ragnar Juelsrud[2] Plamen Nenov[3] Olav Syrstad[4]

Abstract

We study the cross-section of safe asset returns using a tractable asset pricing model with multiple safe assets, agent heterogeneity, transaction costs, and aggregate risk. Changes in the supply or in the transaction costs of a single safe asset induce purifying/polluting effects on the convenience yields of all assets via an aggregate price of convenience. An increase in aggregate risk or risk aversion in our model ends up decreasing liquidity premia via a safety value channel – a repricing effect on risk-free (safe) assets due to a flight to safety adjustment in agents' portfolios. We test the predictions of our model using data on US Treasury yields and changes in Treasury supply. Consistent with our theory, we show that the convenience yield defined as the difference between a maturity matched Treasury-OIS spread and the 3-months Treasury-OIS spread increases with the supply of long maturity bonds and decreases with the supply of shorter maturity bonds. It also increases with the MOVE Index, which is closely correlated with illiquidity and transaction costs in the Treasury market. However, it decreases with the VIX, consistent with our safety value channel. Overall, our tractable model can be useful for analysing the asset pricing effects of central bank market operations as well as unconventional monetary policies.

[1]This paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank. We thank Marie Hoerova and Alp Simsek for useful comments and suggestions.

[2]Norges Bank

[3]Norges Bank and BI Norwegian Business School

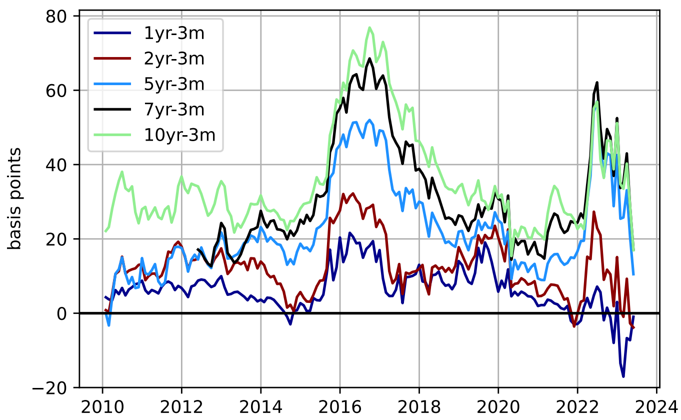
[4]BI Norwegian Business School

4.1 Introduction

Safe assets, such as government bonds and bank deposits, are not just risk-free assets. Investors tend to hold safe assets also for the liquidity services they provide, since such assets are either with a short enough maturity to match investors' liquidity needs, or they can be readily converted for cash in liquid secondary markets with little price impact. Therefore, safe assets tend to command an additional premium, often referred to as a liquidity premium or convenience yield (Krishnamurthy and Vissing-Jorgensen (2012), Brunnermeier et al. (2022)).

However, there is substantial variation within safe assets in terms of their convenience yields. In general, shorter maturity Treasury bills tend to command a larger premium than longer maturity Treasury bonds. Figure 4.1 illustrates this fact by plotting the spread between Treasuries and overnight indexed swaps (OIS) (maturity matched) relative to the 3-months Treasury-OIS spread.[1] As we argue below, this difference in spreads (a double difference) captures the convenience premium of short vs. long maturity US Treasuries. Therefore, longer maturity US Treasuries tend to have a higher discount, or lower convenience premium compared to shorter maturity US Treasuries. This is a general phenomenon that goes beyond US Treasuries. Short maturity safe assets are generally considered the most "money-like" and command the largest premium.

Figure 4.1: Relative Treasury-OIS spread



Notes: This figure shows the US Treasury-overnight indexed swaps (OIS) spread across different maturities relative to the 3-months Treasury-OIS spread. Source: Bloomberg and authors calculations.

Motivated by the variation in convenience premia across safe assets evident in figure 4.1, we

[1]The OIS is a derivative, which is available with different maturities. Two parties swap a fixed rate for a floating overnight rate. It is (close to) risk-free but does embed a term premium.

study in this paper the cross-section of safe asset returns using a highly tractable asset pricing model with multiple safe assets, agent heterogeneity, transaction costs, and aggregate risk. The model is stylised but rich enough to accommodate both a demand for convenience (liquidity), as well as a standard portfolio choice by investors. We use the model to characterise the equilibrium interactions across safe assets due to asset supply effects, as well as the interactions between aggregate risk and the demand for convenience. We test the predictions of our theory using data on US Treasury yields and changes in Treasury supply, for example due to exogenous debt ceiling periods.

In our model agents with standard recursive preferences face uninsurable idiosyncratic risk, against which they self-insure by holding and trading assets. Assets differ both in terms of their transaction costs when traded, which are assumed to be proportional to the value of assets sold, as well as in their exposure to an aggregate risk factor. We refer to the transaction cost on an asset as that asset's illiquidity. A larger transaction cost incurred during sale implies greater illiquidity. We assume there is a menu of risk-free assets that differ in terms of their transaction costs and relative supply, as well as a single risky asset (a market portfolio), which also carries a transaction cost. For tractability, in some of our theoretical analysis, we follow Iachan et al. (2021) and impose a log-linear approximation on preferences and the risky asset's payoffs à-la Campbell and Viceira (2002). This allows us to characterise equilibrium asset returns in closed-form.[2]

In our model the intensity of idiosyncratic shocks that agents are exposed to, combined with their preferences and the whole asset structure, determine the equilibrium liquidity scarcity in this economy and from it an equilibrium price of liquidity or convenience. The aggregate price of convenience is an equilibrium object that reflects the equilibrium liquidity scarcity in this economy relative to a full liquidity insurance first-best, where the marginal rates of substitution between agents with high vs. low interim liquidity needs are equalised. It is, therefore, both a pricing factor, which drives the cross section of asset returns, as well as a sufficient statistic for deviations of this economy from a full insurance first-best.

We refer to the return of a safe asset with positive transaction costs, relative to the most liquid (zero transaction cost) safe asset in the economy as the *convenience yield* of the most liquid safe asset *relative to that asset.* This convenience yield depends on the transaction cost of the asset but also on the aggregate price of convenience in the economy.

Using this framework, we show how changes in the supply or in the transaction cost (i.e. illiquidity) of a single safe asset affects the convenience yields on *all* safe assets via general equi-

[2]This approximation is only relevant for some of our results, particularly related to the effects of aggregate risk.

librium effects operating through the aggregate price of convenience. In our framework, changes to the *composition* of safe asset supply impacts the convenience yields on all safe assets. For example, substituting less liquid with more liquid safe assets one-for-one decreases the aggregate price of convenience and shrinks the convenience yields on *all* safe assets. We refer to this effect as a "purifying (resp. polluting for the opposite substitution) effect". This feature of our model can be used to understand the asset pricing (and welfare) effects of central bank interventions, such as Operation Twist or more recently Quantitative Easing (QE) programs. In our framework, such interventions impact asset prices and welfare in the economy via the aggregate price of convenience. It can also be used to understand the effects of changes in the supply of a single safe asset. In that case, in addition to compositional effects, there is also an effect on the total supply of safe assets. Specifically, in a partial corollary to the purifying/polluting effect, we show that increasing the supply of a safe asset may increase or decrease the convenience yields on all other safe assets, depending on the illiquidity of that safe asset. For example, increasing the supply of a safe asset that is more (less) illiquid than an appropriately-defined average illiquidity in the economy increases (resp. decreases) the convenience yields of *all* safe assets, since it increases the aggregate price of convenience. Therefore, the type of safe asset added to an economy actually matters for aggregate liquidity, and as we also show for welfare.

Finally, we show that when safe assets are more liquid on average than the risky market portfolio – a natural assumption – increases in aggregate risk or risk aversion affect the convenience yields and the price of convenience via a *safety value channel*, which ends up *decreasing* the convenience yields on all safe assets. The main mechanism for this effect is a valuation effect on risk-free assets from higher aggregate risk due to a "flight to safety" adjustment in agents' portfolios. Intuitively, such higher equilibrium valuation of risk-free safe assets relaxes the liquidity needs of agents in this economy and improves their ability to self-insure against idiosyncratic shocks.

Next, we test the predictions of our model using data on US Treasury yields and changes in Treasury supply. In the data we take the difference in Treasury-OIS spreads between longer maturity Treasuries and 3-months Treasury bills (a double difference) as the counterpart to the convenience yield between two safe assets in our model. The first differencing of the OIS rate, which like Treasuries embeds a compensation required for bearing the risk of changes in the short term interest rate over the contract period, i.e. the term premium, allows us to have a cleaner measurement of the convenience yield in Treasuries. The second differencing gives us a measure of the convenience yield in line with our theoretical framework.[3]

[3] The second differencing also takes out any maturity invariant asset type-specific component from the measured convenience yield.

To test for the safe asset supply prediction of our model, we use two sources of variation. First, we use changes in the positions of primary dealers as a proxy for the supply of Treasuries of different maturities to market participants. Intuitively, higher inventory of Treasuries of primary dealers is an indication of large supply and difficulties for dealers to re-sell the Treasuries to other market participants. Quantitatively, a 100 billion increase in the short (long) maturity positions of primary dealers are associated with a decrease (increase) in the convenience premia of the 3-months Treasury bill against other Treasuries with longer maturity of 10 (20) basis point(s) across the entire cross-section. Therefore, consistent with the purifying/polluting effect in our theory, we show that the convenience yield increases (i.e. there is a larger difference between long and short term Treasury-OIS spreads) with the supply of long maturity bonds and decreases with the supply of shorter maturity bonds. Second, we use debt ceiling episodes to get plausibly exogenous variation in the supply of Treasuries. A one percentage point increase in the growth of total outstanding government debt (driven by the debt-ceiling episodes) increases the convenience premia (against all maturities) by 7 basis points. As this is primarily driven by longer maturity Treasuries, this is consistent with the polluting effect of safe asset supply in our model.

Finally, we show that the convenience premia increase with the MOVE Index, which measures yield volatility of US Treasuries, and is thus closely correlated with illiquidity and transaction costs in the Treasury market (Duffie et al. (2023)). However, they actually *decrease* with the VIX, consistent with our safety value channel. Moreover, the effects of interest rate volatility (MOVE Index) and general risk sentiment (VIX) are substantially stronger for Treasuries with long maturities.

Literature review Our paper contributes to the large and growing literature on safe assets and the convenience yield spurred by the seminal contribution of Krishnamurthy and Vissing-Jorgensen (2012) (see Gorton et al. (2012), Caballero and Farhi (2013), Krishnamurthy and Vissing-Jorgensen (2015), Sunderam (2015), Caballero et al. (2016), Nagel (2016), Azzimonti and Yared (2019), He et al. (2019), Gorton (2020), Ahnert and Macchiavelli (2021), Christensen et al. (2021), Diamond and Van Tassel (2021), Jiang et al. (2021), Kacperczyk et al. (2021), Acharya and Dogra (2022), Barro et al. (2022), Brunnermeier et al. (2022), Gorton and Ordonez (2022), Engel and Wu (2023), Krishnamurthy and Li (2023), among others).[4] Krishnamurthy and Vissing-Jorgensen (2012) examine differences in several spreads, including Treasury-corporate bond yield spreads, and corporate bond spreads (adjusted for default risk)

[4]See Brunnermeier and Haddad (2014), Caballero et al. (2017), Golec and Perotti (2017), and Gorton (2017) for a review of the literature.

and examine how those respond to changes in Treasury supply. They use a reduced-form money-in-the-utility model to analyse the convenience services of safe assets, splitting those services into a liquidity and a safety component. An increase in the overall supply of safe assets in their model, decreases the marginal value of convenience provided by each asset. We innovate relative to this seminal paper by explicitly treating different safe assets as imperfect substitutes for liquidity services and examine the resulting cross section of safe-asset returns and its response to the composition of safe assets. While we do not have an explicit difference between liquidity and safety services in our model, it can nonetheless accommodate a safety channel under the natural assumption that safe assets are collectively more liquid than the risky asset in our model economy. Moreover, in an extended version of our model that includes liquidity risk, the resulting differences in assets that are more vs. less exposed to liquidity risk can be interpreted as capturing a safety service.

Our emphasis on transaction costs for understanding the cross-section of safe asset returns relates our paper to Brunnermeier et al. (2022) who emphasise a retrading perspective on safe assets – the ability to retrade safe assets at a low transaction cost and at a predictable value whenever needed. Specifically, the aggregate price of convenience in our framework is conceptually related to the price of service flows in Brunnermeier et al. (2022). Unlike them we have a finite horizon framework, and hence our model cannot feature any rational asset bubbles. Relative to their model we provide a tractable framework that can accommodate a potentially large set of safe assets (in addition to a risky market portfolio), which can be used to understand the effects of changes to the composition of safe assets in an economy – a margin of adjustment that is arguably of first-order importance for monetary policy implementation and understanding the effects of unconventional monetary policy.

Our theoretical model brings our paper close to models of equilibrium asset prices with transaction costs and trading frictions (Amihud and Mendelson (1986), Aiyagari and Gertler (1991), Heaton and Lucas (1996), Vayanos (1998), Vayanos and Vila (1999), Holmström and Tirole (2001), Huang (2003), Lo et al. (2004), Vayanos (2004), Acharya and Pedersen (2005), Duffie et al. (2005)).[5] As in our framework, in these models, transaction costs imply a liquidity premium for assets with lower transaction costs relative to assets with higher transaction costs.[6] Similar to our model, Vayanos and Vila (1999) consider a model with two riskless assets, one

[5]See Amihud et al. (2006) for a review of the literature. See also Constantinides (1986), Duffie and Sun (1990), Davis and Norman (1990), Grossman and Laroque (1990), Dumas and Luciano (1991) for partial equilibrium portfolio allocation models in the presence of transaction costs.

[6]In addition to differences in transaction costs or liquidity across assets, the liquidity premium may arise due to liquidity risk. We also provide a microfoundation for the transaction costs arising due to a specific form of liquidity risk – a correlation between the aggregate liquidity conditions in the economy and an asset's returns, similar to Holmström and Tirole (2001) and Acharya and Pedersen (2005).

without a transaction cost and one with transaction costs. They consider an OLG framework with 3-period-lived agents that trade the assets for life-cycle reasons. Specifically, there are clientele effects resulting in long horizon (agents in the 1st period of life) and short horizon (agents in the 2nd period of life) agents demanding different types of assets. They show that when transaction costs for the illiquid asset increase, its price falls relative to the price of the liquid asset but that the price of the liquid asset also increases. Therefore, the price of the illiquid asset may decrease or increase, depending on the strength of the price response of the liquid asset. If the relative supply of the illiquid asset is large, the strength of the price response is also large, and hence the price of the illiquid asset increases. Similar cross-asset effects of changes in transaction costs are present in our framework as well. Specifically, the liquidity premium in our framework depends on both the relative and overall supply of different asset types. Relative to their model we propose a framework that is tractable enough to accommodate multiple risk-free assets in addition to a risky asset (the market portfolio). Moreover, our model features uninsured idiosyncratic liquidity shocks which affect the resulting convenience yields due to transaction costs beyond the direct effect of transaction costs on returns. In fact it is because of this demand for liquidity that the convenience yield depends on the supply of different asset types.

The explicit treatment of safe assets as imperfect substitutes due to different transaction costs relates our model to a literature on monetary policy implementation that goes back to Tobin (1958) and Tobin (1961) and the portfolio balance channel of monetary policy (Tobin (1969)), as well as a more recent literature on the price effects of the relative supply of different safe assets (Greenwood and Vayanos (2014), Krishnamurthy and Li (2023)) and of large scale asset purchases and QE (Greenwood and Vayanos (2010), Alon et al. (2011), Krishnamurthy and Vissing-Jorgensen (2011), Gagnon et al. (2011), Swanson (2011), Hamilton and Wu (2012)).[7] Our results on the substitution of more for less liquid safe assets are directly related to the portfolio balance channel and also to the "liquidity channel" of QE emphasised by Krishnamurthy and Vissing-Jorgensen (2011), who argue that QE substituted more (central bank reserves) for less liquid assets (US Treasuries), thus increasing the liquidity available in the market and decreasing the aggregate price of convenience and the convenience yield on the most liquid assets relative to less liquid assets. This feature of QE is a direct consequence of the purifying/polluting effect in our model.[8]

[7] See also Brunner and Meltzer (1972), and Andres et al. (2004).

[8] Our model can also easily accommodate a "market functioning channel" if one assumes that transaction costs for an asset are an increasing function of the supply of that asset. See Gagnon et al. (2011) for a discussion of the market functioning channel and also the recent work by Duffie (e.g. Duffie (2020), Duffie and Keane (2023), Duffie et al. (2023)).

Even though we do not explicitly model short vs. long term safe assets in our baseline framework our extended model which features liquidity risk can be interpreted as a theory of the term-structure and the term-premium. In a highly influential paper Vayanos and Vila (2021) model the term-structure of government bonds based on "preferred-habitat" investors as in Modigliani and Sutch (1966) and risk-averse arbitrageurs, which creates partial segmentation of the bond market. Limits to arbitrage imply an underreaction of long rates to short rate shocks and also an increasing term structure, which compensates arbitrageurs for taking risk. Without local demand shocks, changes to demand/supply for a particular bond have "global effects" on all bond prices because of the increase in risk exposure of the arbitrageurs with the effect increasing in bond duration.[9] Conceptually, our framework complements Vayanos and Vila (2021) by also featuring "global effects" of changes in the supply of one particular bond/safe asset on the prices of all safe assets. However, unlike their framework, in our model this occurs via the general equilibrium response of the aggregate price of liquidity.

In our empirical measure of the convenience premium we follow a recent wave of literature applying the OIS-rate as a proxy for the risk-free rate and the corresponding Treasury-OIS spread as a proxy for the price of convenience embedded in Treasuries (Filipovic and Trolle (2013), He et al. (2022), Klingler and Sundaresan (2023), Fleckenstein and Longstaff (2023), Du et al. (2023)). In contrast to the above-mentioned articles, our primary interest is the relative convenience premium across the term structure rather than the level itself. To this end, we deduct the 3-months Treasury-OIS spread from the corresponding longer-term yield spreads.

Du et al. (2023) and Klingler and Sundaresan (2023) relate the increase in the Treasury-OIS spread since the Global Financial Crisis (GFC) to the increase in the supply of Treasuries. Specifically, Du et al. (2023) argue that the Treasury supply is a significant driver of the Treasury-OIS spreads and as a result contributing to the regime shift where dealers have positive net positions in Treasuries. We utilise this finding and use net primary dealer positions as a proxy for excess Treasury supply. Our focus on Treasury supply is also related to Greenwood and Vayanos (2014). The authors investigate the slope of the yield curve by comparing the supply of long vs. short maturity Treasuries (relative supply effects) using a maturity-weighted-debt-to-GDP measure. We also consider such relative supply effects. However, instead of the price of duration risk, we emphasise the price of liquidity and different exposure of Treasuries of different maturities to that price. Also, rather than explicitly looking at the slope of the yield curve we look at the slope of the Treasury-OIS spreads.

[9] Other equilibrium asset pricing models of the term structure include Wachter (2006), Buraschi and Jiltsov (2007), Xiong and Yan (2010), Lettau and Wachter (2011), Gabaix (2012), among others.

4.2 Theory

In this section we present our theoretical framework. We start by informally discussing the main features of our model. Agents in our model have standard Epstein-Zin (Epstein and Zin (1989)) preferences and form portfolios over risk-free and risky assets at an initial period. They may then have interim trading needs, which arise due to uninsurable idiosyncratic shocks as in the asset-pricing with uninsured idiosyncratic risk literature (e.g. Bewley (1979), Aiyagari and Gertler (1991), Constantinides and Duffie (1996), Heaton and Lucas (1996), Heaton and Lucas (2000), Constantinides (2002), Di Tella (2020), Brunnermeier et al. (2022)). In our framework, these idiosyncratic shocks will take the form of shocks to marginal utility, making the agent more impatient to consume in the period. Assets in our framework will each have a proportional transaction cost that is borne by the seller of the asset in the interim period. As in Acharya and Pedersen (2005), we interpret this transaction cost much more broadly than simple fees or bid-ask spreads and instead let this cost encompass liquidity considerations related to price impact and trading delays. In addition, in the appendix we provide a microfoundation where the transaction cost also incorporates liquidity risk – the covariance between the price of an asset and the aggregate liquidity needs in the economy. Finally, we assume that there is always a risk-free asset with zero transaction cost. This is consistent with theories of security design, in which a riskless security has the lowest transaction cost (see Gorton and Pennacchi (1990), DeMarzo and Duffie (1999), Dang et al. (2009)).[10]

4.2.1 Model set-up

The model has 3 periods: $t = 0, 1, 2$. The state at $t = 2$ is uncertain and described by the realisation z of a random variable $Z \sim \text{Lognormal}(\mu, \sigma^2)$. There is a unit measure of 3-period lived investors. Investors have non-storable endowments Y_0 at $t = 0$ and Y_1 at $t = 1$. A measure λ of investors are subject to a liquidity/impatience shock at the beginning of $t = 1$. They are denoted by type $s = \overline{S}$ and are referred to as impatient. The remaining investors are denoted by $s = \underline{S}$ and are referred to as patient. Investors consume in all three periods, $c_0, c_{1,s}, c_{2,s}$. The liquidity shock makes investors value consumption at $t = 1$ more and not value consumption at $t = 2$ anymore. The endowment is the numéraire good in the economy.

There are two types of assets. The first type is risk-free assets, and there is a variety of N

[10] Note that our framework only models the demand for convenience via liquidity needs, and we do not explicitly model any additional demand for safety, for example due to limited participation in markets for risky assets (Vissing-Jorgensen (2002)) or because of special properties of using safe and informationally-insensitive assets as collateral (Dang et al. (2010)). Our model can be easily modified to accommodate the latter channel as borrowing against collateral for liquidity purposes (i.e. funding liquidity) and liquidating an asset (i.e. market liquidity) are closely related concepts.

such assets. We denote the set of these assets by $\mathbb{I} = \{1, 2, ..., N\}$. The second type is risky assets, and for simplicity we assume there is only one such asset, which we refer to as the market portfolio and denote it by m.

For the risk-free assets the payoff is equal to 1 in each state z at $t = 2$. For the market portfolio, we denote the payoff by $\varphi^m(z)$. The total supply of all risk-free assets is denoted by Q^f, while for the market portfolio the total supply is Q^m. The supply of each risk-free asset i is Q^i, and therefore, $Q^f = \sum_{i\in\mathbb{I}} Q^i$. To simplify the analysis, we switch off revaluation/wealth effects and assume that the assets are initially held by one-period lived agents who sell the assets, consume and exit the economy.

Since all investors are *ex ante* symmetric at $t = 0$, we denote the quantity of risk-free asset $i \in \mathbb{I}$ held by an investor at the end of $t = 0$ by X_1^i, while the quantity of the market portfolio is denoted by X_1^m. The quantity of risk-free asset i held by an investor type $s \in \left\{\underline{S}, \overline{S}\right\}$ at the end of $t = 1$ is $X_{2,s}^i$ and analogously for the market portfolio holdings. The prices of the risk-free assets are denoted by P_t^i, and the price of the risky asset by P_t^m. Investors who sell an asset at $t = 1$ pay a proportional transaction cost τ^i for risk-free asset i and τ^m for the market portfolio. We assume that for risk-free asset $i = 1$, $\tau^1 = 0$. Therefore, there is one risk-free asset without transaction costs in this economy. Lastly, investors are assumed to have recursive Epstein-Zin (Epstein and Zin, 1989) preferences given by:

$$U_0 = \log(c_0) + \chi_s \log(c_{1,s}) + \beta_s \log(U_{2,s}),$$

where $U_{2,s} = \left(E\left[c_2^{1-\gamma}\right]\right)^{1/(1-\gamma)}$, and γ measures the degree of risk aversion. To model the $t = 1$ liquidity shock of the investors, we assume that

$$\chi_s = \begin{cases} \underline{\chi}, & s = \underline{S} \\ \overline{\chi}, & s = \overline{S} \end{cases}, \text{ with } 1 = \underline{\chi} < \overline{\chi}, \text{ and } \beta_s = \begin{cases} 1, & s = \underline{S} \\ 0, & s = \overline{S}. \end{cases}$$

In what follows we focus on the case where $\overline{\chi} = 2$. This is a convenient parametrisation, since given the number of periods and the assumption of no discounting between periods, it implies that there are no total utility differences between the patient and impatient investors, which generates tractability in our mathematical expressions below at the cost of little loss in generality. We can write the $t = 0$ preferences of an investor of type s recursively as

$$U_0 = \log(c_0) + U_{1,s}, \tag{4.1}$$

with $U_{1,s} = \chi_s \log(c_{1,s}) + \beta_s \log(U_{2,s})$ and $U_{2,s} = \left(E\left[c_{2,s}^{(1-\gamma)}\right]\right)^{1/(1-\gamma)}$. At $t = 1$ we have

$$U_{1,s} = \begin{cases} 2\log(c_{1,\overline{S}}), & s = \overline{S} \\ \log(c_{1,\underline{S}}) + \log(U_{2,\underline{S}}), & s = \underline{S}. \end{cases} \tag{4.2}$$

The $t = 0$ budget constraint of a representative investor is given by

$$c_0 + \sum_{i=1}^{N} P_0^i X_1^i + P_0^m X_1^m = Y_0. \tag{4.3}$$

We conjecture that all risk-free assets in period 1 have the same price P_1^f and verify this conjecture in the appendix (see section 4.6.1). Therefore, the $t = 1$ budget constraint for a type-s investor is given by

$$\begin{aligned} c_{1,s} + A_{2,s} = Y_1 + A_1 - \sum_{i \in \mathbb{I}} \tau^i P_1^f \max\left\{0, X_{1,s}^i - X_{2,s}^i\right\} \\ + \tau^m P_1^m \max\left\{0, X_{1,s}^m - X_{2,s}^m\right\}, \end{aligned} \tag{4.4}$$

where the last two terms on the right-hand side reflects the transaction costs incurred from selling assets, and A_1 denotes the asset holdings the investor has at the beginning of $t = 1$, namely $A_1 \equiv P_1^f X_{1,s}^f + P_1^m X_{1,s}^m$, with $X_{1,s}^f \equiv \sum_{i \in \mathbb{I}} X_{1,s}^i$. $A_{2,s}$ denotes the asset holdings of the investor at the end of $t = 1$, with $A_{2,\underline{S}} \equiv P_1^f X_{2,\underline{S}}^f + P_1^m X_{2,\underline{S}}^m$, where $X_{2,s}^f \equiv \sum_{i \in \mathbb{I}} X_{2,s}^i$. Finally, the period 2 budget constraint of a type-s investor in state z is given by

$$c_{2,s}(z) = X_{2,s}^f + \varphi^m(z) X_{2,s}^m. \tag{4.5}$$

Next, we define an equilibrium for this economy.

Definition 4 (**Equilibrium**). *An equilibrium for this economy consists of period 0 asset and consumption choices of the investors, $\left\{X_1^i\right\}_{i \in \mathbb{I}}$, X_1^m, and c_0; period 1 asset and period 1 and 2 consumption choices for type-s investors, $\left\{X_{2,s}^i\right\}_{i \in \mathbb{I}}$, $X_{2,s}^2$, $c_{1,s}$, and $c_{2,s}(z)$; and period 0 and 1 asset prices, $\left\{P_0^i\right\}_i$, P_0^m, P_1^f, P_1^m, such that*

a) given prices, the consumption allocations, and asset holdings solve the period 0 and period 1 problems of the investor (i.e. Eq. (4.1) subject to the period $t = 0$ budget constraint in Eq. (4.3), as well as Eq. (4.2) subject to the period $t = 1$ budget constraint in Eq. (4.4), as well as the period 2 budget constraint in Eq. (4.5) which must hold for all z);

b) given the period 0 and period 1 asset holdings of investors, the asset markets clear in both

$t = 0$ and $t = 1$, that is

$$
\begin{aligned}
X_1^i &= Q^i, \ \forall i \in \mathbb{I} \\
X_1^m &= Q^m
\end{aligned}
$$

and

$$
\begin{aligned}
X_{2,\overline{S}}^i + X_{2,\underline{S}}^i &= Q^i, \ \forall i \in \mathbb{I} \\
X_{2,\overline{S}}^m + X_{2,\underline{S}}^m &= Q^m.
\end{aligned}
$$

4.2.2 Characterisation

We solve the model backwards in time. We first look at the $t = 1$ problems and market clearing conditions given the asset holdings of the investors as of the beginning of $t = 1$ and then move to the $t = 0$ problem and market clearing conditions given some anticipated $t = 1$ prices. Imposing rational expectations and given the $t = 0$ choices of agents this gives a full equilibrium characterisation of this economy.

4.2.2.1 $t = 1$ characterisation

From $t = 1$ onwards, there are two types of investors: patient and impatient. An impatient investor sells all her assets because she derives no utility from consumption in future periods. A patient investor's problem can be split into a consumption-saving problem and a portfolio choice problem.[11] In the consumption-saving problem, she decides how much to consume in the current period and how much to save for the next period. In the portfolio choice problem, she decides how to allocate her savings over assets. Specifically, below we denote by w the portfolio share invested in the market portfolio. Market clearing implies that patient investors hold all assets at the end of $t = 1$. The following lemma summarises the $t = 1$ characterisation.

Lemma 1 ($t = 1$ **characterisation**). *Let $A_1 \equiv \sum_{i \in \mathbb{I}} P_1^f Q_1^i + P_1^m Q_1^m$ denote the equilibrium value of financial wealth at $t = 1$. The following characterises the equilibrium prices and allocations at $t = 1$ and $t = 2$.*

a) Investors' $t = 1$ consumption and saving decisions:

$$
\begin{aligned}
c_{1,\overline{S}} &= \frac{2}{1+\lambda} Y_1 - \sum_{i \in \mathbb{I}} \tau^i Q^i - \tau^m Q^m \\
c_{1,\underline{S}} &= \frac{1}{1+\lambda} Y_1
\end{aligned}
$$

[11] This is implied by the assertion that in equilibrium she only increases her asset holdings at $t = 1$, as we verify in the appendix section 4.6.1.1.

and

$$A_{2,\overline{S}} = 0$$
$$A_{2,\underline{S}} = \frac{Y_1 + A_1}{2}.$$

b) Patient investors' portfolio allocations solve:

$$U_2(A_{2,\underline{S}}) = \max_{X^f_{2,\underline{S}}, X^m_{2,\underline{S}}} \left(E\left[c_{2,\underline{S}}(z)^{(1-\gamma)} \right] \right)^{1/(1-\gamma)}$$

s.t.

$$P^f_1 X^f_{2,\underline{S}} + P^m_1 X^m_{2,\underline{S}} = A_{2,\underline{S}}$$
$$c_{2,\underline{S}}(z) = X^f_{2,\underline{S}} + \varphi^m(z) X^m_{2,\underline{S}},\ \forall z.$$

c) Prices are determined by:

$$\frac{1}{(1-\lambda)} P^f_1 Q^f = (1-w)\frac{Y_1 + A_1}{2}$$

$$\frac{1}{(1-\lambda)} P^m_1 Q^m = w\frac{Y_1 + A_1}{2}$$

where $w \equiv \frac{P^m_1 X^m_{2,\underline{S}}}{A_{2,\underline{S}}}$, $0 \leq w \leq 1$, *is the portfolio share invested by patient investors in the market portfolio. Moreover,* $P^i_1 = P^f_1\ \forall i \in \mathbb{I}$.

Proof. See appendix 4.6.1. □

4.2.2.2 $t = 0$ characterisation

At $t = 0$, investors choose their current consumption and their asset portfolio. The problem is characterised by a standard Euler equation. The marginal utility of today's consumption must equal the expected marginal utility of tomorrow's consumption times the return on each asset earned by holding it from the first to the second period, where the expectation is taken over the idiosyncratic state at $t = 1$. All uninvested resources are consumed, and there is asset market clearing, with all assets held by the investors. Lemma 2 summarises the equilibrium equations that characterise the economy at $t = 0$.

Lemma 2 ($t = 0$ **characterisation**)**.** *Given anticipated period* $t = 1$ *asset prices* $P^i_1, \forall i \in \mathbb{I}$ *and* P^m_1, *the following list of equations characterises the economy at* $t = 0$.

a) Prices are determined by the Euler equations with respect to every asset:

$$\frac{\partial U_0}{\partial c_0} = \lambda \frac{\partial U_{1,\overline{S}}}{\partial c_{1,\overline{S}}} \frac{P_1^j}{P_0^j}(1-\tau^j) + (1-\lambda) \frac{\partial U_{1,\underline{S}}}{\partial c_{1,\underline{S}}} \frac{P_1^j}{P_0^j}, \; \forall j \in \mathbb{I} \cup m.$$

b) The portfolio allocations, or asset holdings, are:

$$X_1^i = Q^i, \; \forall i \in \mathbb{I}$$
$$X_1^m = Q^m.$$

c) Consumption given prices and asset holdings are given by:

$$c_0 = Y_0 - \sum_{i \in \mathbb{I}}^{N} P_0^i Q^i - P_0^m Q^m.$$

Proof. See appendix 4.6.1. □

4.2.3 Approximating the second period portfolio choice

For some of our theoretical results below, we will simplify the $t = 1$ portfolio choice problem of the patient investor by applying a log-normal approximation for the equilibrium asset returns in Iachan et al. (2021), which in turn is based on the log-normal portfolio choice approximation in Campbell and Viceira (2002). Under that approximation, the portfolio choice problem of the patient investor simplifies to a mean-variance portfolio choice problem, given normally distributed log asset returns. To set-up the simplified problem, we define the log returns $r_2^m(z) \equiv \log\left(\frac{\varphi^m(z)}{P_1^m}\right)$ and $r_2^f \equiv \log\left(\frac{1}{P_1^f}\right)$. Also, we define the log of the certainty-equivalent return on the portfolio R_2^{ce} by $r_2^{ce} \equiv \log(R_2^{ce})$ and the risk premium on the risky asset by $\pi \equiv E[r_2^m(z)] + \frac{\sigma^2}{2} - r_2^f$. The portfolio maximisation problem then simplifies to

$$r_2^{ce} - r_2^f = \max_w \; w\pi - \frac{\gamma}{2} w^2 \sigma^2.$$

The first-order condition shows that the investor invests the following constant share into the risky asset:

$$w = \frac{\pi}{\gamma \sigma^2}. \tag{4.6}$$

The certainty equivalent log return of her portfolio is then

$$r_2^{ce} = r_2^f + \frac{1}{2} \frac{\pi^2}{\gamma \sigma^2}. \tag{4.7}$$

This approximation requires a minor adjustment in our equilibrium definition as well. In particular, the investor is assumed to act according to Eqs. (4.6) and (4.7) when forming $t = 1$ portfolios and making her $t = 0$ and $t = 1$ consumption-saving decisions.

We point out that from our main theoretical results in section 4.3 only proposition 10 and the general case in proposition 9 depend on this log-approximation. The other results are independent of it.

4.2.4 Microfoundation of transaction costs

In the appendix section 4.6.1.2, we provide microfoundations for the transaction costs. We first show two examples where we adjust our framework to incorporate assets with different maturities, a feature we are particularly interested in in light of our empirical explorations in section 4.4.2. In both examples we show that long maturity assets are less convenient to accommodate agents' interim liquidity needs and can therefore be associated with carrying higher transaction costs. In the first example the reason for that is that there is a mismatch between the maturity of the long maturity asset and the horizon of liquidity needs. In the second example the $t = 1$ return on the long maturity asset is negatively correlated with the marginal utility of consumption, making it inconvenient to hold.

For both examples we adapt our framework by assuming, for simplicity, that there are only two risk-free assets. We further modify the framework slightly by assuming that one asset matures in period $t = 1$, while the other matures in $t = 2$, which introduces a term structure into our model.

In the first example we additionally assume that if the two-period asset has to be resold at $t = 1$, the investor has to pay a fee, which can be thought of as a market access fee. Our results show that the long maturity asset is less convenient and therefore may be associated with higher transaction costs. Unlike the short maturity asset, which pays off precisely when an investor has liquidity needs, the long maturity asset has to be liquidated, while incurring a cost to do so.

In the second example we do not assume that agents incur any transaction costs when selling assets. In contrast, we focus on the impact of aggregate uncertainty on the convenience of assets with different maturities. As a concrete example (which we later generalise) we assume that the endowment in $t = 1$ is stochastic. There is a news shock about the aggregate endowment realised in $t = 1$, so that at $t = 0$ agents do not know if their endowment will be high or low in the second period. Our results reveal that the price, and hence the return, of the long maturity asset in period $t = 1$ covaries negatively with the marginal utility of consumption. In comparison, the short maturity asset matures in $t = 1$ and therefore its payoff is not correlated with endowment uncertainty. The negative covariance between the future return and future marginal utility (and therefore also with the aggregate price of convenience) is another reason why the long maturity asset is less convenient to be held at $t = 0$ and trades at a discount relative to the short maturity asset.

We also generalise these examples, by deriving a general expression for the convenience yield within our specialised set-up with two assets with different maturities. We assume both market fees and aggregate uncertainty, thus combining the two examples. Compared to the second example we further assume that the aggregate uncertainty can be with respect to any exogenous variable, for example payoffs. Additionally, we assume an isoelastic utility function instead of a log utility function to generalise further. We also assume that the third period is discounted with a discount factor which is possibly smaller than one (to allow for aggregate uncertainty on the discount factor). We show that the expression for the convenience yield nests our two examples from beforehand, and that the drivers are the same. With respect to the influence of aggregate uncertainty it depends on the specific kind of uncertainty if the long or the short maturity asset are more convenient. Lastly, it is important to emphasise that, in line with general asset pricing theory, it is the covariance of returns with aggregate risk and not the variance of aggregate risk that influences equilibrium asset prices and therefore the convenience yield.

4.3 Theoretical results

4.3.1 The price of convenience

An important equilibrium object that is central to our theoretical results is the equilibrium price of convenience. To define this object, let $m_{1,s} \equiv \frac{\partial U_{1,s}}{\partial c_{1,s}} / \frac{\partial U_0}{\partial c_0}$ denote the ratio of marginal utilities of consumption for an investor of type $s \in \{\underline{S}, \overline{S}\}$. Define

$$\eta \equiv \frac{m_{1,\overline{S}}}{m_{1,\underline{S}}} = \frac{\partial U_{1,\overline{S}}}{\partial c_{1,\overline{S}}} / \frac{\partial U_{1,\underline{S}}}{\partial c_{1,\underline{S}}}.$$

Therefore, η capture the degree to which marginal utilities of consumption between patient and impatient investors are aligned in equilibrium. Indeed, as we show in proposition 11 in the appendix, in the first-best full liquidity insurance case, $\eta = 1$ and there is full equalisation of marginal utilities across idiosyncratic states at $t = 1$. Moreover, the higher is η, the further away is this economy from the full-insurance benchmark, so that η is also a sufficient statistic for how far the economy is from the first-best allocation (where $\eta = 1$).

In our decentralised economy generally $\eta \geq 1$, with the value of η determined in equilibrium. Therefore, η can be thought of as reflecting the equilibrium liquidity scarcity in this economy. We therefore refer to η as the *price of convenience*. We can express the price of convenience η as a function of model primitives and asset prices,

$$\eta = \frac{\frac{2}{1+\lambda}Y_1}{\frac{2}{1+\lambda}Y_1 - \widetilde{\tau}P_1^f Q^f - \tau^m P_1^m Q^m},$$

where $\widetilde{\tau} \equiv \sum_{i\in\mathbb{I}}^N \tau^i \frac{Q^i}{Q^f}$ is defined as the average transaction cost of risk-free assets in the economy. Therefore, the price of convenience depends on the period $t = 1$ endowment Y_1, the share of impatient investors λ and the transaction costs incurred in equilibrium, i.e. $\widetilde{\tau}P_1^f Q^f + \tau^m P_1^m Q^m$.[12] In the special case when $Q^m = 0$, η simplifies further to

$$\eta = \frac{1}{1-(1-\lambda)\widetilde{\tau}/2}.$$

In that case η depends on the average transaction costs $\widetilde{\tau}$ and the share of impatient investors λ. Specifically, η increases in $\widetilde{\tau}$. Intuitively, a higher average transaction cost worsens the equilibrium liquidity scarcity and decreases the self-insurance possibilities, leading to a higher value of η. η also decreases with the share of impatient investors, λ. The more impatient investors, the lower the $t = 1$ price of the safe assets due to the larger supply pressure and cash-in-the-market pricing from the patient investors. However, a lower $t = 1$ price actually *mitigates* the impact of the transaction costs (as those are incurred proportional to the value of assets sold), improving liquidity insurance and leading to a lower η.

4.3.2 Convenience yields

Next, we characterise the $t = 0$ convenience yield of an asset in the economy, which we define as the relative return of that asset relative to the return on the most liquid risk-free asset in the economy. Equivalently, this is the liquidity premium of the most liquid safe asset in the economy relative to a less liquid asset. We define the period $t = 0$ gross return on an asset $i \in \mathbb{I} \cup \{m\}$ as $R_1^i \equiv \frac{P_1^i}{P_0^i}$. We state our first main result:

Proposition 6 (**Convenience yields**). *Define* $\psi^j \equiv \frac{R_1^j}{R_1^1}$. *Then*

$$\psi^i = \frac{1}{1-\lambda\tau^i - m_{1,\overline{S}}\tau^i R_1^1} = \frac{1}{1-\tau^i\frac{\lambda}{\lambda+(1-\lambda)/\eta}} \textit{ for all } i \in \mathbb{I} \cup m.$$

[12]As the endowment is non-storable the price of convenience does not depend on the endowment in the first period, Y_0.

Proof. See appendix 4.6.1. □

Proposition 6 shows that the convenience yield ψ^i is influenced by three forces. First, an asset's gross return has to compensate for the transaction cost incurred by selling the asset. Second, an asset's return has to compensate for the transaction cost being incurred precisely when the investor is impatient. Third, how much the transaction cost matters for the convenience yield depends on the price of convenience η.

4.3.3 Effects of transaction costs

We next show how a change in an asset's transaction cost impacts the convenience premia across the entire cross-section of assets in the economy.

Proposition 7 (Transaction costs). *The convenience yield, ψ^i, is increasing in the transaction cost of asset i, τ^i, i.e. $\frac{\partial \psi^i}{\partial \tau^i} > 0$, $\forall i \in \mathbb{I} \cup m$. It is also increasing in the transaction cost of any other asset, τ^k, i.e. $\frac{\partial \psi^i}{\partial \tau^k} > 0$, for $k \neq i$ and k and $i \in \mathbb{I} \cup m$. The impact is higher, the higher is τ^i, i.e. $\frac{\partial^2 \psi^i}{\partial {\tau^i}^2} > 0$ and $\frac{\partial^2 \psi^i}{\partial \tau^k \partial \tau^i} > 0$, $\forall i \in \mathbb{I} \cup m$.*

Proof. See appendix 4.6.1. □

Therefore, a change in an asset's transaction cost not only impacts its own convenience yield but also impacts the entire cross-section of assets. This spillover effect is driven by a change in the price of convenience η. An increase in the transaction cost of any asset *increases* the price of convenience as overall liquidity in the economy decreases. A higher price of convenience affects all assets and raises all the convenience yields, making the most liquid asset more valuable. The more illiquid an asset is in terms of the transaction cost, the more its corresponding convenience yield is affected. Intuitively, if investors have to pay high transaction costs, then they are more exposed to price increases driven by the increase of the price of convenience.

4.3.4 Effects of changes in asset supplies

We next show that changes in the composition and overall supply of safe assets also affect the entire cross-section of convenience yields. We first consider a compositional change in the supply of the risk-free assets due to an increase in the supply of one asset and an equivalent decrease in the supply of another asset.

Proposition 8 (Purifying/polluting effects). *Given a compositional change where $dQ^l = -dQ^k$ for two risk-free assets k and l, the convenience yield ψ^i for any asset $i \in \mathbb{I} \cup m$ increases strictly in response to this change iff $\tau^l > \tau^k$.*

Proof. See appendix 4.6.1. □

Substituting more liquid for less liquid assets one-for-one can be interpreted as a "purification" of the pool of safe assets by altering its composition. This decreases the aggregate price of convenience and shrinks the convenience premia with respect to all assets, or the liquidity premium on the most liquid asset. We call this the "purifying/polluting effect" of safe asset supply.

This result of our theoretical analysis allows us to analyse the asset pricing (and welfare) effects of central bank interventions, such as Operation Twist or more recently QE programs. Since in both cases the central bank substitutes more for less liquid assets, such interventions decrease the convenience yields and increase welfare in the economy (as they lead to a lower value of η, see appendix section 4.6.1.2).

A similar effect occurs when the overall supply of a specific risk-free asset changes. In that case, in addition to the shift in the composition of safe assets, there is an overall increase in the supply of safe assets and an implied shift in the ratio of safe to risky assets. To switch off this second effect, we first assume that there is no risky asset in the economy, $Q^m = 0$. In a second step we also discuss the general case.

Proposition 9 (**Asset supply effect**). *If $Q^m = 0$, then for any risk-free asset l*

$$\frac{\partial \psi^i}{\partial Q^l} > 0 \text{ iff } \tau^l > \tilde{\tau} \text{ for all } i \in \mathbb{I}.$$

If $Q^m > 0$ and $\left(\frac{P_1^f - P_1^m}{P_1^m}\right)$ is close to zero, then for any risk-free asset l

$$\frac{\partial \psi^i}{\partial Q^l} > 0, \text{ for all } i \in \mathbb{I} \cup \{m\}, \text{ iff } \tau^l P_1^f + \tau^m Q^m \frac{\partial P_1^m}{\partial Q^l} + \tilde{\tau} Q^f \frac{\partial P_1^f}{\partial Q^l} > 0.$$

Proof. See appendix 4.6.1. □

The intuition for this result is very similar to that of the purifying/polluting effect from proposition 8. Whether convenience yields increase or decrease now depends on whether the risk-free asset whose supply increases is more or less liquid compared to the average transaction cost on risk-free assets. If it is more (less) liquid, then there is a purifying (polluting) effect from increasing the supply of that asset, and convenience yields decrease (increase). If the supply of the risky asset is positive, then in addition to the purifying/polluting effect across the set of risk-free assets, there are also revaluation effects of the risky and risk-free assets. This additional effect complicates the characterisation but the broad intuition stays the same. Lastly, let us point out that the general result holds under the assumption that the risk premium on the market portfolio is relatively small and therefore $\left(\frac{P_1^f - P_1^m}{P_1^m}\right)$ is close to zero. We view this assumption as empirically justifiable given the measured risk premia on risky assets.

Therefore, the analysis of asset supply effects shows that which safe assets are included in an economy matters for aggregate liquidity conditions, asset pricing and welfare.

4.3.5 Effects of risk and risk premia

Our model features aggregate risk through the risky market portfolio. In light of crisis and flight to safety episodes it is interesting to examine the impact of risk and risk-aversion on the cross-section of safe asset returns according to our framework. The following proposition discusses the effects through the lens of our model.

Proposition 10 (Impact of risk). *If* $\left(\frac{P_1^f - P_1^m}{P_1^m}\right)$ *is close to zero, then, for all* $i \in \mathbb{I} \cup m$, $\frac{\partial \psi^i}{\partial \sigma^2} < 0$, *iff* $\tau^m > \widetilde{\tau}$, *and* $\frac{\partial \psi^i}{\partial \gamma} < 0$, *iff* $\tau^m > \widetilde{\tau}$.

Proof. See appendix 4.6.1. □

An increase in the payoff variance σ^2 or the risk aversion parameter γ increases the price of the risk-free assets and decreases the price of the risky asset. If safe assets are on average more liquid than the risky asset – which we view as the natural assumption – then an increase in aggregate risk or risk aversion implies that this revaluation effect increases the value of safe assets and *improves* their ability to self-insure. Put differently, the price of convenience decreases, and so do the convenience yields for the whole cross-section of safe assets.

4.4 Empirical evidence

We test our theoretical predictions on the cross-section of safe asset returns by examining the US Treasury market. Before showing the various tests, we discuss the data and the actual measurement of convenience yields.

4.4.1 Data and measurement

4.4.1.1 Data sources

We use data on Treasury yields and overnight indexed swap (OIS) rates to measure convenience yields (see section 4.4.1.2). We then add data from different sources, depending on the specific empirical test. To test our theoretical predictions about supply effects, we also use data on primary dealer net positions and total outstanding US government debt. To test the effects of transaction costs and risk, we use data on the MOVE Index and the VIX in the US. The VIX measures stock market volatility, while the MOVE Index measures implied bond volatility in the US Treasury market (Mallick et al., 2017). Further, in some regressions we also add information on the SP 500 Index and the Effective Federal Funds Rate as control variables. Table 4.1 presents summary statistics.[13]

Table 4.1: Summary statistics US

Data	Maturity	Mean	Median	St. Dev.
Treasury-OIS	1Y	7.07	6.29	6.01
spread in bp	2Y	13.10	12.33	8.11
	5Y	22.84	18.97	13.16
	7Y	32.47	24.29	15.41
	10Y	36.34	31.76	14.42
Primary dealer	(0y, 1y]	2.34	1.95	1.6
net positions (defl.)	(1y, 3y]	2.77	2.15	2.66
in 10 bil USD	(3y, 6y]	1.03	1.08	1.39
	(6y, 11y]	0.29	0.50	1.44
Debt growth in %		0.14	0.05	0.28
MOVE Index		73.48	69.30	21.71
VIX		18.49	16.64	7.19
SP 500 Index		2398.87	2124.39	988.37
Effective Fed Funds Rate in %		0.63	0.15	0.85

4.4.1.2 Measuring the convenience yield

We measure the convenience yield ψ^i as the spread between maturity-matched Treasuries and OIS relative to the 3-months Treasury-OIS spread. The OIS is a derivative where two parties swap a fixed rate for a floating overnight rate. Concretely, this means that the receiver of the fixed rate pays the realised overnight rate while the counterpart pays the fixed rate and receives the overnight rate. The OIS is available with different maturities. Since there is no exchange of the notional amount, the OIS rate does not embed any convenience connected to "store of value" or "liquidity services". In addition, a variation margin is mandatory and the OIS rate is therefore (close to) risk free. However, as in Treasuries, the OIS rate *does* embed the term premium, i.e. a compensation required for bearing the risk of changes in the short term interest

[13] The time horizon used for the summary statistics is by default 2010-2022. Only for the primary dealer net positions we "only" use data up to 2021 and for the debt growth up to 2023 as we use this time horizon in the regressions below. For the variables used in several regressions with different time horizons, we use data up to 2022. If the frequency is not weekly, we change it to weekly and use the last value of the week.

rate over the contract period. We therefore argue that the difference between the Treasury rate and maturity matched OIS rate partials out the term premium from the respective Treasury yield, allowing us to have a term-premium free measure of the convenience yield. Furthermore, we get rid of any balance-sheet related matters by looking at the relative Treasury-OIS spread where we deduct the 3-months spread from any other maturity spread. As pointed out by He et al. (2022), balance sheet-related matters were especially salient during COVID and were an important component of convenience yields. The double-differenced measure of the convenience yield also gives us a measure in line with the definition of the convenience yield in our theoretical framework as the relative return of two safe assets.

In what follows, we discuss in detail analytically all the components of the Treasury-OIS spread[14] and why our double-differenced measure of the Treasury-OIS spread at a certain maturity minus the 3-months Treasury-OIS spread provides a model-consistent measure of the convenience yield. Suppose the return on a security consists of the following components:

$$r^{\text{term}} + r^{\text{safety}} + r^{\text{convenience}} + r^{\text{balance sheet}}.$$

The first part r^{term} measures the return one obtains for lending money over a specific term. The other three parts measure any additional return due to payoff risk, liquidity and balance sheet matters. The Treasury-OIS spread (TO), assuming that both have a 5 year maturity for expositional purposes, is therefore given by:

$$\text{TO}_{5Y} = (r^{\text{term}}_{T,\,5Y} + r^{\text{safety}}_{T,\,5Y} + r^{\text{convenience}}_{T,\,5Y} + r^{\text{balance sheet}}_{T,\,5Y}) - (r^{\text{term}}_{OIS,\,5Y} + r^{\text{safety}}_{OIS,\,5Y} + r^{\text{convenience}}_{OIS,\,5Y} + r^{\text{balance sheet}}_{OIS,\,5Y}).$$

As pointed out above, both the Treasury and the OIS are risk-free, $r^{\text{safety}}_{T,\,5Y} = r^{\text{safety}}_{OIS,\,5Y} = 0$ and incorporate the same term premium, $r^{\text{term}}_{T,\,5Y} - r^{\text{term}}_{OIS,\,5Y} = 0$. Therefore the spread simplifies to:

$$\text{TO}_{5Y} = (r^{\text{convenience}}_{T,\,5Y} + r^{\text{balance sheet}}_{T,\,5Y}) - (r^{\text{convenience}}_{OIS,\,5Y} + r^{\text{balance sheet}}_{OIS,\,5Y}).$$

The relative Treasury-OIS spread, where we deduct the 3-months spread from the spread of any maturity (here 5 years), is then given by

$$\begin{aligned}\text{TO}_{5Y} - \text{TO}_{3M} = {} & [(r^{\text{convenience}}_{T,\,5Y} + r^{\text{balance sheet}}_{T,\,5Y}) - (r^{\text{convenience}}_{OIS,\,5Y} + r^{\text{balance sheet}}_{OIS,\,5Y})] \\ & - [(r^{\text{convenience}}_{T,\,3M} + r^{\text{balance sheet}}_{T,\,3M}) - (r^{\text{convenience}}_{OIS,\,3M} + r^{\text{balance sheet}}_{OIS,\,3M})].\end{aligned}$$

We assume that any differences of Treasuries to OIS with respect to balance sheet matters do

[14] Figure 4.4 in the appendix plots the Treasury-OIS spreads.

not depend on the maturity, $r_{T,\,5Y}^{\text{balance sheet}} - r_{OIS,\,5Y}^{\text{balance sheet}} = r_{T,\,3M}^{\text{balance sheet}} - r_{OIS,\,3M}^{\text{balance sheet}}$. Above we discussed that the OIS does not give any liquidity services, but even if it would embed any liquidity premia, we assume that they are independent of the maturity, $r_{OIS,\,5Y}^{\text{convenience}} = r_{OIS,\,3M}^{\text{convenience}}$. Then

$$\mathrm{TO}_{5Y} - \mathrm{TO}_{3M} = r_{T,\,5Y}^{\text{convenience}} - r_{T,\,3M}^{\text{convenience}}.$$

Therefore our double difference in the Treasury-OIS spread measures the convenience yield.

In our microfoundation of the transaction costs, see section 4.2.4, we discussed that longer maturity assets potentially have higher transaction costs and therefore carry a lower convenience yield. Figure 4.1 reveals that this is indeed the case. Longer maturity safe assets tend to command a higher discount, or lower convenience premium compared to shorter maturity safe assets. This is a general phenomenon, not limited to Treasuries, as short maturity safe assets are generally considered the most "money-like".

4.4.2 Evidence

4.4.2.1 Supply effects

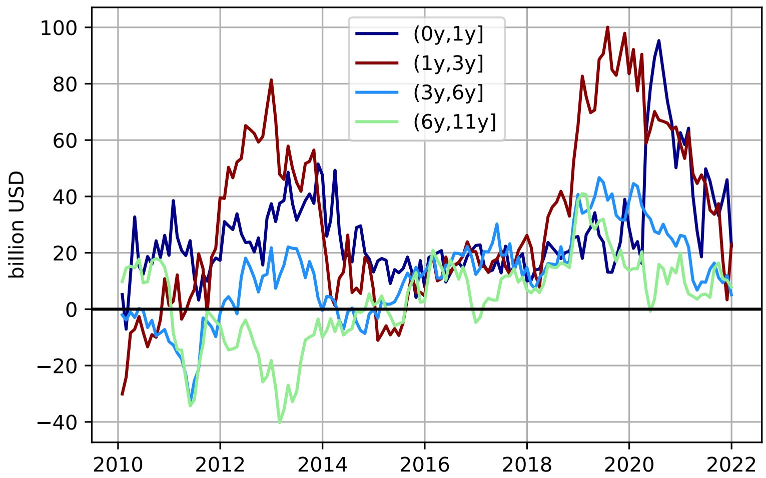
First, we test our theoretical prediction of the aggregate supply effect described in proposition 9. The proposition states that if we increase the total supply of an asset with higher (lower) than average transaction costs, then this will lead to an increase (decrease) in convenience yields with respect to the *entire* cross-section of risk-free assets. We referred to this increase in convenience yields implied by an increase in the supply of less liquid assets with above-average transaction costs, as a polluting effect of safe asset supply (see section 4.3.3).

Variation in supply: primary dealer net positions As a first source of variation to test this prediction, we use primary dealers' net positions in Treasuries. They are a proxy for the supply of Treasuries of different maturities to market participants. Higher primary dealer holdings of Treasuries may indicate a large available supply in the market. One reason why their inventories can grow large is that primary dealers have to place bids in the Treasury auctions even in the absence of customer interest.[15] Figure 4.2 shows the evolution of primary dealers net positions of Treasuries, and document substantial variation across time and across maturities.

We regress the convenience yields, denoted by $\mathrm{TO}_{M,t} - \mathrm{TO}_{3M,t}$, on the net positions of primary dealers (OLS regression). The outcome variable consists of the convenience yields with respect to all maturities (1y, 2y, 5y, 7y, 10y), denoted by M. Each maturity at each date t is a separate observation. The explanatory variables are the deflated primary dealers'

[15] See for example Fleming et al. (2024) on how primary dealers manage their inventories.

Figure 4.2: Primary dealer net positions in Treasuries



Notes: This figure shows the primary dealer net positions across different maturities. Source: NY FED Primary Dealer Statistics and authors calculations.

net positions in Treasuries above and below the volume-weighted average maturity separately. Primary dealers' net positions are divided into four maturity baskets: (0y, 1y], (1y, 3y], (3y, 6y], (6y, 11y]. The average volume-weighted maturity is given by 2y.[16] Therefore, we classify the first two as short maturity baskets and denote their sum in period t as "positions$_t$ with $\tau < \tilde{\tau}$". We classify the other two as long maturity baskets and call their sum in period t "positions$_t$ with $\tau > \tilde{\tau}$".[17] In an additional regression we also control for the SP 500 Index[18] and the Effective Federal Funds Rate. We also add fixed effects for all maturities M. The frequency of our data is weekly and the time horizon is 2010-2021.[19] The regression equation (2) is given by

$$\begin{aligned} \mathrm{TO}_{M,t} - \mathrm{TO}_{3M,t} = \alpha_M &+ \beta_1 \left(\text{Positions}_t \text{ with } \tau < \tilde{\tau}\right) + \beta_2 \left(\text{Positions}_t \text{ with } \tau > \tilde{\tau}\right) \\ &+ \beta_3 \ln(\text{SP } 500_t) + \beta_4 \text{ Effective Federal Funds Rate}_t + e_t \end{aligned}$$

The results are shown in table 4.2. Quantitatively, an increase of 100 billion in the short (long) maturity positions of primary dealers corresponds to a decrease (increase) in the convenience yields of around 10 (20) basis point(s) across the entire cross-section (columns 1 and 2). This is consistent with our theory. Convenience yields increase with the supply of long

[16]The average maturity is calculated as 0.5*volume (0y, 1y] basket+1.5*volume (1y, 3y] basket+4.5*volume (3y, 6y] basket+8.5*volume (6y, 11y] basket. The weights are the average maturities of the baskets.

[17]The bonds with maturities above 2y and below 3y are only about 1.5% of the (1y, 3y] basket. They are reported separately from the rest of the basket after 2013, but not before.

[18]Consistent with the price of convenience and our welfare discussion, this variable can indicate how agents in the economy value liquidity.

[19]We use the data up to 2021 because the data in the (0y, 1y] basket might be misreported for 2022.

maturity bonds and decrease with the supply of short maturity bonds. The coefficients are highly significant. With the additional control variables and maturity fixed effects (column 2), R^2 increases significantly, otherwise the results are similar. In appendix 4.6.4 we show that our results also hold when we run the regression for each convenience yield maturity individually.

Table 4.2: Impact of supply fluctuations (measured by net positions)

	(1) TO_M-TO_{3M}	(2) TO_M-TO_{3M}
Positions with $\tau < \tilde{\tau}$	-1.1022*** (0.195)	-1.0748*** (0.139)
Positions with $\tau > \tilde{\tau}$	1.7880*** (0.244)	2.1084*** (0.319)
ln(SP 500)		-3.8990*** (1.393)
Effective Federal Funds Rate		-0.4955 (0.717)
Constant	24.3204*** (1.367)	40.0322*** (10.762)
Maturity FE	No	Yes
N	3004	2959
R^2	0.092	0.556

Notes: The outcome variable consists of the convenience yields of all maturities M (1y, 2y, 5y, 7y, 10y), denoted by $TO_{M,t} - TO_{3M,t}$. The explanatory variables are the deflated primary dealers' net positions in Treasuries above and below the volume weighted average maturity separately, the SP 500 Index and the Effective Federal Funds Rate. Fixed effects (FE) are added for every maturity M. The frequency is weekly and the time horizon is 2010-2021. Standard errors are heteroscedasticity and autocorrelation robust (HAC) using 6 lags and without small sample correction. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

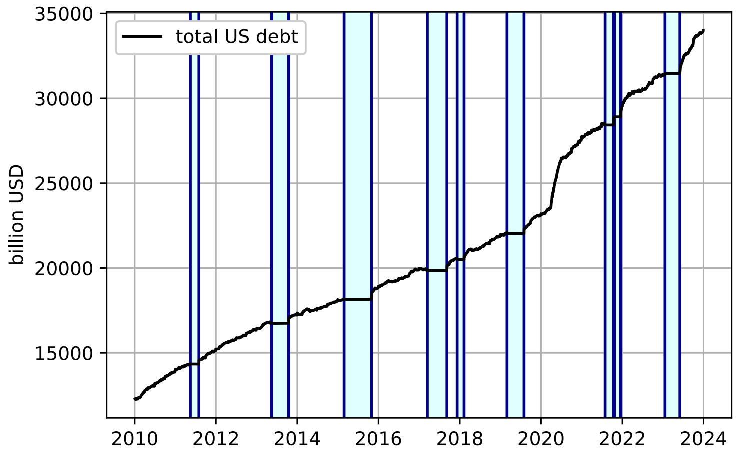
Variation in supply: debt ceiling dates As a second measure of supply variation, we use exogenous debt ceiling dates.[20] In the first stage, we regress weekly US debt growth (calculated from total outstanding US government debt) on a dummy variable called "Debt ceiling dates":

$$\text{Debt growth}_t = \alpha + \beta \text{ Debt ceiling dates}_t + u_t.$$

[20]See Cashin et al. (2017) for a study of how debt ceilings affect Treasury yields.

The dummy variable equals one in all episodes in which the daily total outstanding debt did not grow over a period of at least five weeks (i.e., debt ceilings were in place).[21] We interpret these episodes as a negative shock to longer-term debt, as debt growth is mostly driven by long maturities. Total bills outstanding are a small fraction of total debt and grow comparatively slowly (see figure 4.5 in the appendix). We use our variable debt ceiling dates as the instrument as the debt ceiling periods are exogenous.

Figure 4.3: Constant episodes of at least 5 weeks



Notes: This figure shows the daily total outstanding US government debt. All episodes where the debt growth was constant for at least five weeks are marked. Source: Bloomberg and authors calculations.

In the second stage, the outcome variable consists of the convenience premia of all maturities (1y, 2y, 5y, 7y, 10y), denoted by $\mathrm{TO}_{M,t} - \mathrm{TO}_{3M,t}$, as before. Each maturity at each date t is a separate observation. The frequency is weekly and the time horizon is 2010-2023. The explanatory variable consists of the fitted values of debt growth. This means that debt growth only affects convenience discounts through debt ceiling dates. Therefore, the second stage is as follows:

$$\mathrm{TO}_{M,t} - \mathrm{TO}_{3M,t} = \alpha + \beta^{2SLS}\, \widehat{\mathrm{Debt}}\ \mathrm{growth}_t + e_t.$$

As shown in table 4.3 and consistent with our theory, a one percentage point increase in the growth of government debt (driven by the debt ceiling episodes) increases the convenience yields (with respect to all maturities) by 7 basis points.[22] In appendix 4.6.4 we show that our

[21]No growth is defined if the debt does not change by more than 10 billion from day to day. As we can see in figure 4.3 this seems to be an accurate measure.

[22]The first-stage shows that our definition of debt ceilings (no change over a period of at least 5 weeks) is appropriate.

results also hold when we run the regression for each convenience yield maturity individually.

In sum, the findings in this section highlight that the supply of bonds affect convenience yields in the direction predicted by the model. Next, we move on to explore the role of transaction costs and risk for understanding convenience yields.

Table 4.3: Impact of supply fluctuations (measured by debt ceilings)

	Debt growth	TO_M-TO_{3M}
Debt ceiling dates	-0.1710***	
	(0.008)	
Debt growth		7.4238***
		(3.7222)
Constant	0.1736***	20.919***
	(0.008)	(0.5479)
Model	First-stage	IV-2SLS
N	3364	3364
F-Statistic	433.6	

Notes: The outcome variable consists of the convenience yields of all maturities M (1y, 2y, 5y, 7y, 10y), denoted by $TO_{M,t} - TO_{3M,t}$. The explanatory variable consists of the fitted values of debt growth. The instrument are all days on which daily debt growth was constant for at least five weeks. The frequency is weekly and the time horizon is 2010-2023. Standard errors are heteroscedasticity robust. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

4.4.2.2 Effects of transaction costs and risk

Finally, we test the theoretical predictions of the propositions 7 and 10. The first proposition states that an increase in the transaction costs of *any* of the assets increases convenience premia with respect to the *entire* cross-section. The effect on convenience premia is stronger for assets with higher transaction costs. The second proposition states that an increase in the volatility of risky assets or risk aversion reduces the convenience returns of the fully liquid safe asset against the *entire* cross-section of safe assets.

We again use an OLS regression to test our theoretical predictions. As in the other regressions, the outcome variable consists of the convenience yields of all maturities (1y, 2y, 5y, 7y, 10y), denoted by $TO_{M,t} - TO_{3M,t}$ as before. Each maturity at each date t is a separate observation. The explanatory variables are the MOVE Index and the VIX. The MOVE Index measures the yield volatility of Treasuries, which is closely correlated with illiquidity and transaction costs in the Treasury market (Duffie et al. (2023)).[23] It is therefore an appropriate measure to test our theoretical prediction. As a measure of risk sentiment, we use the VIX.

We also control for the SP 500 Index and the Effective Federal Funds Rate as well as fixed effects for each maturity M in a separate regression. The frequency is weekly and the time horizon is 2010-2022. We also add regressions with a dummy term D indicating that the convenience yields are with respect to short maturities (1y and 2y). The regression equation (4) is given by

$$\begin{aligned}\mathrm{TO}_{M,t} - \mathrm{TO3}_{M,t} = \alpha_M &+ \beta_1 \ln(\mathrm{MOVE}_t) + \beta_2\, \mathcal{I}_{m\leq 2} \ln(\mathrm{MOVE}_t)\\ &+ \beta_3 \ln(\mathrm{VIX}_t) + \beta_4\, \mathcal{I}_{m\leq 2} \ln(\mathrm{VIX}_t)\\ &+ \beta_5 \ln(\mathrm{SP\ 500}_t) + \beta_6 \text{ Effective Federal Funds Rate}_t + e_t.\end{aligned}$$

The results are shown in table 4.4. The convenience yield correlates positively with the MOVE Index and negatively with the VIX, consistent with our theory. A 1 percentage point increase in the MOVE Index (VIX) is associated with an increase (decrease) in the convenience yield by 10 (9) basis points (column 1). In appendix 4.6.4 we show that our results also hold when we run the regression for each convenience yield maturity individually. The effects of interest rate volatility (MOVE Index) and general risk sentiment (VIX) are much stronger for long maturity Treasuries (column 3).[24] Adding the additional control variables and fixed effects to the regression does not change the results qualitatively (columns 2 and 4).

[23] In our model we interpret transaction costs broader than bid-ask spreads. Therefore we consider the MOVE Index as an appropriate measure.

[24] Our theory also predicts this with respect to increases in the MOVE Index. We do not give any prediction with respect to the VIX.

Table 4.4: Impact of the MOVE Index and the VIX

	(1) TO_M-TO_{3M}	(2) TO_M-TO_{3M}	(3) TO_M-TO_{3M}	(4) TO_M-TO_{3M}
ln(MOVE)	9.8316***	11.2478***	15.9234***	16.7004***
	(2.441)	(1.701)	(1.920)	(2.225)
D*ln(MOVE)			-12.9009***	-13.4007***
			(1.661)	(2.716)
ln(VIX)	-9.2804***	-8.7156***	-13.9169***	-13.6295***
	(1.870)	(1.314)	(2.066)	(1.855)
D*ln(VIX)			12.0794***	12.0384***
			(2.334)	(2.064)
ln(SP 500)		1.9879**		1.9790**
		(1.120)		(1.094)
Effective Federal Funds Rate		2.2247***		2.2149***
		(0.492)		(0.463)
Constant	6.6740	-32.6248***	2.4758	-19.0647
	(10.546)	(12.833)	(6.923)	(12.834)
Maturity FE	No	Yes	No	Yes
N	3263	3263	3263	3263
R^2	0.036	0.528	0.433	0.544

Notes: The outcome variable consists of the convenience yields of all maturities M (1y, 2y, 5y, 7y, 10y), denoted by $\mathrm{TO}_{M,t} - \mathrm{TO}_{3M,t}$. The explanatory variables are the MOVE Index and the VIX, the SP 500 Index and the Effective Federal Funds Rate. Fixed effects (FE) are added for every maturity M. The dummy term D indicates that the convenience premia are with respect to short maturities (1y and 2y). The frequency is weekly and the time horizon is 2010-2022. Standard errors are heteroscedasticity and autocorrelation robust (HAC) using 6 lags and without small sample correction. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

4.5 Conclusion

This paper presents an equilibrium theory of the cross section of the convenience yields across asset returns. It is based on transaction costs and investors' demand for self-insurance against idiosyncratic liquidity needs. Central to our theory is the price of convenience, an equilibrium object, which is also a sufficient statistic for deviations from the first-best full insurance case.

We characterise the liquidity yield in our model and show how it depends on the transaction cost of an asset as well as the price of convenience. We then show that there is a purifying (resp. polluting) effect from rebalancing safe asset supply when increasing (decreasing) the supply of more liquid assets. Changes in risk or risk aversion also impact convenience yields in our model via changes in the price of convenience due to repricing effects. Our model is supported by a variety of new facts that we document about the US Treasury market.

The fact that our decentralised equilibrium does not achieve the first best allocation is intriguing and points to pecuniary externalities that matter for aggregate welfare in the presence of transaction costs. Examining further these potential channels of inefficiency would be important for any future research on this topic.

4.6 Appendix

4.6.1 Omitted proofs

4.6.1.1 Proof of lemma 1

Proof. We describe and solve the optimization problems in the second and third period.

$t = 1$ **problem of patient investors** For the problem of the patient investor in period 1, we split the problem into a portfolio choice problem and a consumption-saving problem. Specifically, we conjecture that in equilibrium she only increases her position in each asset. We then split her problem in two: a consumption-saving decision (problem 1) given by

$$\begin{aligned} U_{1,\underline{S}} = \max_{c_{1,\underline{S}}, A_{2,\underline{S}}} & \log(c_{1,\underline{S}}) + \log(U_2(A_{2,\underline{S}})) \\ \text{s.t.} & \\ c_{1,\underline{S}} + A_{2,\underline{S}} = Y_1 + A_1, & \end{aligned}$$

where $A_{2,\underline{S}}$ denotes saving into $t = 2$ and a portfolio choice problem (problem 2), given by

$$\begin{aligned} U_2(A_{2,\underline{S}}) = \max_{X^f_{2,\underline{S}}, X^m_{2,\underline{S}}} & \left(E\left[c_{2,\underline{S}}\left(z\right)^{(1-\gamma)} \right] \right)^{1/(1-\gamma)} \\ \text{s.t.} & \\ P^f_1 X^f_{2,\underline{S}} + P^m_1 X^m_{2,\underline{S}} = A_{2,\underline{S}} & \\ c_{2,\underline{S}}\left(z\right) = & X^f_{2,\underline{S}} + \varphi^m\left(z\right) X^m_{2,\underline{S}}, \ \forall z. \end{aligned}$$

To write down the portfolio choice problem we used $P^f_1 = P^i_1$. We demonstrate in a separate proof below why this is the case.

Next, we discuss the consumption-saving problem. Given homothetic preferences and fully capitalizable income, we show in lemma 4 below that the value function $U_2(A_{2,\underline{S}})$ is linear in savings, i.e.

$$U_2(A_{2,\underline{S}}) = R^{CE}_2 A_{2,\underline{S}}.$$

The consumption-saving problem in period 1 becomes

$$\begin{aligned} U_{1,\underline{S}} = \max_{c_{1,\underline{S}}, A_{2,\underline{S}}} & \log(c_{1,\underline{S}}) + \log(R^{CE}_2 A_{2,\underline{S}}) \\ \text{s.t.} & \\ c_{1,\underline{S}} + A_{2,\underline{S}} = Y_1 + A_1. & \end{aligned}$$

With log utility, we then have $A_{2,\underline{S}} = \frac{Y_1+A_1}{2}$ and $c_{1,\underline{S}} = \frac{Y_1+A_1}{2}$. We insert the optimal values into the utility function $U_{1,\underline{S}}$ and derive the value function, which we denote by $V_{1,\underline{S}}$. It is given by $V_{1,\underline{S}} = \log R_2^{CE} \left(\frac{Y_1+A_1}{2}\right)^2$.

$t = 1$ **problem of impatient investors** The problem of the impatient investor in period 1 is trivial as she sells all her asset holdings and consumes all available resources at $t = 1$. Therefore

$$X^i_{2,\overline{S}} = 0 \; \forall i$$
$$X^m_{2,\overline{S}} = 0$$

and we have $c_{1,\overline{S}} = Y_1 + A_1 - \sum_i \tau_1^i P_1^f Q^i - \tau^m P_1^m Q^m$, where $A_1 \equiv \sum_i P_1^f Q^i + P_1^m Q^m$ denotes the investor's financial wealth at $t = 1$, excluding transaction costs. The investor does not build up any savings for period 2, $A_{2,\overline{S}} = 0$ and does not consume in this period, $c_{2,\overline{S}}(z) = 0$.

Market clearing at $t = 1$ At $t = 1$ the patient and impatient investors trade assets with one another. Taking into account the respective mass of each type of investor and the asset holding decisions of impatient investors, market clearing conditions are given by

$$(1-\lambda)X^i_{2,\underline{S}} = Q^i \; \forall i \in \mathbb{I}$$
$$(1-\lambda)X^m_{2,\underline{S}} = Q^m.$$

Therefore, patient investors hold all the assets at the end of $t = 1$. From these market clearing conditions we can derive the period 1 prices as well as savings and consumption. Using the market clearing condition $X^m_{2,\underline{S}} = \frac{1}{(1-\lambda)} Q^m$, $A_{2,\underline{S}} = \frac{Y_1+A_1}{2}$ from the consumption-saving decision and the portfolio weight w on the risky asset, we end up with the following equation:

$$\frac{1}{(1-\lambda)} P_1^m Q^m = w \frac{Y_1 + A_1}{2}.$$

For the risk-free assets, we use the market clearing combined with the optimal portfolio allocation to get

$$\frac{1}{(1-\lambda)} P_1^f Q^f = (1-w) \frac{Y_1 + A_1}{2}.$$

Summing the two above equations yields the savings

$$A_1 = \frac{(1-\lambda)}{(1+\lambda)} Y_1.$$

Lastly, the budget constraints in combination with the optimal asset holdings yield the consumption choices in all periods:

$$c_{1,\overline{S}} = \frac{2}{1+\lambda}Y_1 - \sum_i \tau^i P_1^f Q^i - \tau^m P_1^m Q^m$$
$$c_{1,\underline{S}} = \frac{1}{1+\lambda}Y_1.$$

$\square$

4.6.1.2 Proof of lemma 2

Proof. We describe and solve the optimization problem in the first period.

$t = 0$ problem of investors The period $t = 0$ problem of the representative investor is given by

$$U_0 = \max_{c_0, \{X_1^i\}_{i\in\mathbb{I}}, X_1^m} \log(c_0) + \lambda V_{1,\overline{S}} + (1-\lambda)V_{1,\underline{S}}$$
$$\text{s.t.}$$
$$c_0 = Y_0 - \sum_{i\in\mathbb{I}} P_0^i X_1^i - P_0^m X_1^m,$$

where $V_{1,\overline{S}} = 2\log(c_{1,\overline{S}})$ with $c_{1,\overline{S}} = Y_1 + A_1 - \sum_{i\in\mathbb{I}} \tau^i P_1^i X_1^i - \tau^m P_1^m X_1^m$ and $V_{1,\underline{S}} = \log R_2^{CE}(c_{1,\underline{S}})^2$ with $c_{1,\underline{S}} = \frac{Y_1+A_1}{2}$. The Euler equation for holdings of asset $i \in \mathbb{I} \cup \{m\}$ is

$$\frac{\partial U_0}{\partial c_0} = E_s\left[\frac{\partial V_{1,s}}{\partial A_1}\frac{P_1^i}{P_0^i}(1 - \mathcal{I}_{s=\overline{S}}\tau^i)\right],$$

or

$$\frac{\partial U_0}{\partial c_0} = \lambda\frac{\partial U_{1,\overline{S}}}{\partial c_{1,\overline{S}}}\frac{P_1^i}{P_0^i}(1-\tau^i) + (1-\lambda)\frac{\partial U_{1,\underline{S}}}{\partial c_{1,\underline{S}}}\frac{P_1^i}{P_0^i},$$

where we have used the fact that at $t = 1$, by the envelope theorem for the agent's $t = 1$ problem, $\frac{\partial V_{1,s}}{\partial A_1} = \frac{\partial U_{1,s}}{\partial c_{1,s}}$. Therefore, at the optimum, the marginal utility of consumption today equals the expected marginal utility of consumption tomorrow, where the expectation is taken over the realisation of the idiosyncratic investor's state s and the transaction cost incurred in the impatient state is taken into account.

Market clearing at $t = 0$ Market clearing at $t = 0$ is given by

$$X_1^i = Q^i, \ \forall i$$
$$X_1^m = Q^m.$$

Additionally, we prove the following two lemmas:

Lemma 3 (Second period prices). $P_1^i = P_1^f, \forall i \in \mathbb{I}$.

The Lagrangian of the portfolio choice problem of the patient investor in period 1 is given by

$$\mathcal{L} = \left(\int_z \left[(R_2^f X_{2,\underline{S}}^f + R_2^m(z) X_{2,\underline{S}}^m)^{(1-\gamma)} \right] d(z) \right)^{\frac{1}{(1-\gamma)}}$$
$$+ \mu \left[A_{2,\underline{S}} - \sum_i P_1^i X_{2,\underline{S}}^i - P_1^m X_{2,\underline{S}}^m \right].$$

The first order conditions with respect to the risk-free assets (given an interior solution) are:

$$\frac{1}{(1-\gamma)} \left(\int_z \left[(R_2^f X_{2,\underline{S}}^f + R_2^m(z) X_{2,\underline{S}}^m)^{(1-\gamma)} \right] dF(z) \right)^{\frac{1}{(1-\gamma)}-1}$$
$$= \frac{\mu P_1^i}{(1-\gamma) R_2^f \left[\int_z (R_2^f X_{2,\underline{S}}^f + R_2^m(z) X_{2,\underline{S}}^m)^{(-\gamma)} dF(z) \right]}$$

for all $i \in \mathbb{I}$. It follows that P_1^i is the same for all $i \in \mathbb{I}$. We denote this price by P_1^f.

Lemma 4 (Third period utility). $U_2(A_{2,\underline{S}}) = R_2^{CE} A_{2,\underline{S}}$.

Using $P_i^1 = P_f^1$, for all i, the portfolio choice problem of the patient investor in period 1 can be rewritten as

$$\max_{X_{2,\underline{S}}^m} \left[\int_z \left[\frac{R_2^f}{P_1^f} A_{2,\underline{S}} + \left(\frac{R_2^m(z)}{P_1^m} - \frac{R_2^f}{P_1^f} \right) P_1^m X_{2,\underline{S}}^m \right]^{(1-\gamma)} dF(z) \right]^{\frac{1}{(1-\gamma)}}.$$

The first order condition is given by

$$\int_z \left[\frac{R_2^f}{P_1^f} A_{2,\underline{S}} + \left(\frac{R_2^m(z)}{P_1^m} - \frac{R_2^f}{P_1^f} \right) P_1^m X_{2,\underline{S}}^m \right]^{(-\gamma)} \left[\left(\frac{R_2^m(z)}{P_1^m} - \frac{R_2^f}{P_1^f} \right) P_1^m \right] dF(z) = 0.$$

We guess that $X^m_{2,\underline{S}} = \theta A_{2,\underline{S}}$ and insert it:

$$\int\limits_z \left[\frac{R^f_2}{P^f_1} A_{2,\underline{S}} + \left(\frac{R^m_2(z)}{P^m_1} - \frac{R^f_2}{P^f_1} \right) P^m_1 \theta A_{2,\underline{S}} \right]^{(-\gamma)} \left[\left(\frac{R^m_2(z)}{P^m_1} - \frac{R^f_2}{P^f_1} \right) P^m_1 \right] dF(z) = 0.$$

It follows that θ does not depend on $A_{2,\underline{S}}$ and is determined by the following equation:

$$\int\limits_z \left[\frac{R^f_2}{P^f_1} + \left(\frac{R^m_2(z)}{P^m_1} - \frac{R^f_2}{P^f_1} \right) P^m_1 \theta \right]^{(-\gamma)} \left[\left(\frac{R^m_2(z)}{P^m_1} - \frac{R^f_2}{P^f_1} \right) P^m_1 \right] dF(z) = 0.$$

We insert $X^m_{2,\underline{S}} = \theta A_{2,\underline{S}}$ into the maximisation problem. This yields

$$\max_{X^m_{2,\underline{S}}} \left[\int\limits_z \left[\frac{R^f_2}{P^f_1} A_{2,\underline{S}} + \left(\frac{R^m_2(z)}{P^m_1} - \frac{R^f_2}{P^f_1} \right) P^m_1 \theta A_{2,\underline{S}} \right]^{(1-\gamma)} dF(z) \right]^{\frac{1}{(1-\gamma)}},$$

or

$$\max_{A_{2,\underline{S}}} R^{CE}_2 A_{2,\underline{S}}$$

where $R^{CE}_2 \equiv \left[\int\limits_z \left[\frac{R^f_2}{P^f_1} + \left(\frac{R^m_2(z)}{P^m_1} - \frac{R^f_2}{P^f_1} \right) P^m_1 \theta \right]^{(1-\gamma)} dF(z) \right]^{\frac{1}{(1-\gamma)}}$. Note that this is the certainty equivalent return R^{CE}_2 without applying any approximation as we do in section 4.2.3. $\square$

Proof of proposition 6

Proof. We know that the first order conditions of the period $t = 0$ problem for all assets are

$$\frac{1}{c_0} = E_s \left(\frac{\partial U_{1,s}}{\partial c_{1,s}} \frac{P^f_1 (1 - \mathcal{I}_{s=\overline{S}} \tau^i)}{P^i_0} \right)$$

for all $i \in \mathbb{I} \cup m$. Let us define $R^i_{1,s} \equiv \frac{P^f_1 (1 - \mathcal{I}_{s=\overline{S}} \tau^i)}{P^i_0}$, $R^1_1 \equiv \frac{P^f_1}{P^1_0}$, and $m_{1,s} \equiv \frac{\frac{\partial U_{1,s}}{\partial c_{1,s}}}{\frac{\partial U_0}{\partial c_0}}$. Then $\frac{\partial U_0}{\partial c_0} = E_s \left(\frac{\partial U_{1,s}}{\partial c_{1,s}} R^i_{1,s} \right)$ for all i. This means that also $\frac{\partial U_0}{\partial c_0} = R^1_1 E_s \left(\frac{\partial U_{1,s}}{\partial c_{1,s}} \right)$ must hold. We rearrange the general formula as follows:

$$
\begin{aligned}
0 &= E_s\left(\frac{\partial U_{1,s}}{\partial c_{1,s}} R^i_{1,s} - \frac{\partial U_0}{\partial c_0}\right)\\
0 &= E_s\left(\frac{\partial U_{1,s}}{\partial c_{1,s}}\left(R^i_{1,s} - R^1_1\right)\right)\\
0 &= E_s\left(\frac{\partial U_{1,s}}{\partial c_{1,s}}\right) E_s\left(R^i_{1,s} - R^1_1\right) + COV\left(\frac{\partial U_{1,s}}{\partial c_{1,s}}, \left(R^i_{1,s} - R^1_1\right)\right)\\
-\frac{COV\left(\frac{\partial U_{1,s}}{\partial c_{1,s}}, \left(R^i_{1,s} - R^1_1\right)\right)}{E_s\left(\frac{\partial U_{1,s}}{\partial c_{1,s}}\right)} &= E_s\left(R^i_{1,s}\right) - R^1_1\\
-COV\left(m_s, \left(R^i_{1,s} - R^1_1\right)\right) R^1_1 &= E_s\left(R^i_{1,s}\right) - R^1_1\\
-COV\left(m_s, R^i_{1,s}\right) &= \frac{E_s\left(R^i_{1,s}\right)}{R^1_1} - 1.
\end{aligned}
$$

We define the convenience yield $\psi^i \equiv \frac{R^i_1}{R^1_1}$ with $R^i_1 \equiv \frac{P^f_1}{P^i_0}$. We insert for $R^i_{1,s}$ and rearrange:

$$
\begin{aligned}
\frac{E_s\left(\frac{P^f_1(1-\mathcal{I}_{s=\overline{S}}\tau^i)}{P^i_0}\right)}{R^1_1} - 1 &= -COV\left(m_{1,s}, \frac{P^f_1(1-\mathcal{I}_{s=\overline{S}}\tau^i)}{P^i_0}\right)\\
\frac{\frac{P^f_1}{P^i_0}}{R^1_1}(1-\lambda\tau^i) - 1 &= -\frac{\frac{P^f_1}{P^i_0}}{R^1_1} COV\left(m_{1,s}, (1-\mathcal{I}_{s=\overline{S}}\tau^i)\right) R^1_1\\
\psi^i &= \frac{1}{(1-\lambda\tau^i) - COV\left(m_{1,s}, \mathcal{I}_{s=\overline{S}}\tau^i\right) R^1_1}.
\end{aligned}
$$

We then rearrange the covariance term:

$$
\begin{aligned}
COV\left(m_{1,s}, \mathcal{I}_{s=\overline{S}}\tau^i\right) &= \lambda\tau^i m_{\overline{S}} - E_s\left(m_s\right)\lambda\tau^i\\
COV\left(m_{1,s}, \mathcal{I}_{s=\overline{S}}\tau^i\right) &= \lambda\tau^i\left(m_{\overline{S}} - \frac{1}{R^1_1}\right).
\end{aligned}
$$

We insert it into our main term of interest:

$$
\begin{aligned}
\psi^i &= \frac{1}{1-\lambda\tau^i m_{\overline{S}} R^1_1}\\
\psi^i &= \frac{1}{1-\lambda\tau^i m_{\overline{S}} \frac{1}{E_s(m_s)}}\\
\psi^i &= \frac{1}{1-\tau^i \frac{\lambda m_{\overline{S}}}{\lambda m_{\overline{S}} + (1-\lambda) m_{\underline{S}}}}.
\end{aligned}
$$

Therefore the convenience yield is given by

$$
\psi^i = \frac{1}{1-\tau^i \frac{\lambda}{\lambda + (1-\lambda)/\eta}}
$$

where $\eta \equiv \frac{m_{\overline{S}}}{m_{\underline{S}}}$ is the relative kernel, which we later call price of convenience. □

Proof of proposition 7

Proof. We analyse the impact of the fees on the convenience yields. The derivatives are given by

$$\frac{\partial \psi^i}{\partial \tau^i} = -\left(\psi^i\right)^2 \frac{\lambda}{\lambda + (1-\lambda)/\eta} \left(\frac{(1-\lambda)\tau^i}{\lambda + (1-\lambda)/\eta} \frac{\partial (1/\eta)}{\partial \tau^i} - 1 \right) > 0$$

and

$$\frac{\partial \psi^i}{\partial \tau^k} = \frac{\partial \psi^i}{\partial (1/\eta)} \frac{\partial (1/\eta)}{\partial \tau^k} > 0 \text{ for all } i \neq k,$$

where $\frac{\partial \psi^i}{\partial (1/\eta)} = -\left(\psi^i\right)^2 \frac{\tau^i \lambda (1-\lambda)}{(\lambda + (1-\lambda)/\eta)^2} < 0$, $\frac{\partial (1/\eta)}{\partial \tau^i} = \frac{-P_1^i Q^i}{\frac{2}{1+\lambda} Y_1} < 0$, and $\frac{\partial (1/\eta)}{\partial \tau^k} = \frac{-P_1^k Q^k}{\frac{2}{1+\lambda} Y_1} < 0$ and i and $k \in \mathbb{I} \cup m$. In addition it follows that

$$\begin{aligned}
\frac{\partial^2 \psi^i}{\partial {\tau^i}^2} =& -2\psi^i \frac{\partial \psi^i}{\partial \tau^i} \frac{\lambda}{\lambda + (1-\lambda)/\eta} \left(\frac{(1-\lambda)\tau^i}{\lambda + (1-\lambda)/\eta} \frac{\partial (1/\eta)}{\partial \tau^i} - 1 \right) \\
& - \left(\psi^i\right)^2 \frac{\lambda}{\lambda + (1-\lambda)/\eta} \left(\frac{(1-\lambda)}{\lambda + (1-\lambda)/\eta} \frac{\partial (1/\eta)}{\partial \tau^i} - 1 \right) > 0 \\
\frac{\partial^2 \psi^i}{\partial \tau^k \partial \tau^i} =& -\left(2\psi^j \frac{\partial \psi^i}{\partial \tau^i} \frac{\tau^i \lambda (1-\lambda)}{(\lambda + (1-\lambda)/\eta)^2} + \left(\psi^i\right)^2 \frac{\lambda (1-\lambda)}{(\lambda + (1-\lambda)/\eta)^2} \right) \frac{\partial (1/\eta)}{\partial \tau^k} > 0.
\end{aligned}$$

□

Proof of propositon 8

Proof. We study the impact on the convenience yields due to a compositional change in the supply of the risk-free assets. Suppose $dQ^l = -dQ^k$ for any l and k in $\in \mathbb{I}$. Then

$$\frac{d\psi^i}{dQ^l} = \frac{\partial \psi^i}{\partial (1/\eta)} \frac{d(1/\eta)}{dQ^l}$$

where $\frac{\partial \psi^i}{\partial (1/\eta)} = -\left(\psi^i\right)^2 \frac{\tau^i \lambda (1-\lambda)}{(\lambda + (1-\lambda)/\eta)^2} < 0$, $\frac{d(1/\eta)}{dQ^l} = \frac{\partial (1/\eta)}{\partial Q^l} + \frac{\partial (1/\eta)}{\partial Q^k} \frac{dQ^k}{dQ^l} = \left(\tau^k - \tau^l\right) \frac{1}{\frac{2}{1+\lambda} Y_1}$ for all $i \in \mathbb{I} \cup m$. □

Proof of proposition 9

Proof. We analyse the impact of an overall supply change for two cases: The case where there is no risky asset in the economy, i.e. $Q^m = 0$ and where there is, i.e. $Q^m > 0$. For the first case, the following lemma applies:

Lemma 5 (No risk). *If $Q^m = 0$, then $\eta = \frac{2}{2-(1-\lambda)\bar{\tau}}$.*

We will use this simplified price of convenience to analyse the first case. An overall change in the supply of the risk-free assets due to an increase in the supply of one of the assets, suppose asset $l \in \mathbb{I}$, is given by

$$\frac{\partial \psi^i}{\partial Q^l} = \frac{\partial \psi^i}{\partial (1/\eta)} \frac{\partial (1/\eta)}{\partial Q^l}$$

where $\frac{\partial \psi^i}{\partial(1/\eta)} = -\left(\psi^i\right)^2 \frac{\tau^i \lambda(1-\lambda)}{(\lambda+(1-\lambda)/\eta)^2} < 0$ for all $i \in \mathbb{I} \cup m$. $\frac{\partial(1/\eta)}{\partial Q^l}$ depends on which case we look at. If $Q^m = 0$, then $\frac{\partial(1/\eta)}{\partial Q^l} = \frac{(1-\lambda)}{2}\left(\widetilde{\tau} - \tau^l\right)\frac{1}{Q^1}$. For the second case we assume that $\left(\frac{P_1^f - P_1^m}{P_1^m}\right)$ is close to zero. This allows for a closed-form solution for prices (see section 4.6.1.2). If in this case $Q^m > 0$, then

$$\frac{\partial(1/\eta)}{\partial Q^l} = -\frac{1}{\frac{2}{1+\lambda}Y_1}\left[\tau^l P_1^f + \widetilde{\tau} Q^f \frac{\partial P_1^f}{\partial Q^l} + \tau^m Q^m \frac{\partial P_1^m}{\partial Q^l}\right]$$

with

$$\begin{aligned}
\frac{\partial P_1^f}{\partial Q^l} &= -\frac{1}{Q^f}(1-\lambda)\frac{1}{1+\lambda}Y_1\left[\frac{1}{Q^f} - \frac{1}{\gamma\sigma_{22}}\left(\frac{\pi}{Q^f} - \frac{\partial \pi}{\partial Q^f}\right)\right] \\
\frac{\partial P_1^m}{\partial Q^l} &= \frac{1}{Q^m}(1-\lambda)\frac{1}{1+\lambda}Y_1\left[\frac{1}{\gamma\sigma_{22}}\frac{\partial \pi}{\partial Q^f}\right]
\end{aligned}$$

and

$$\frac{\partial \pi}{\partial Q^f} = \frac{\partial \mathcal{A}}{\partial Q^f} + \frac{1}{2}\left(\mathcal{A}^2 + \frac{Q^m}{Q^f}\gamma\sigma_{22}\right)^{-\frac{1}{2}}\left(\mathcal{A} - \gamma\sigma_{22}\right)\frac{Q^m}{\left[Q^f\right]^2} > 0 \text{ if } \mathcal{A} > \gamma\sigma_{22}.$$

□

Proof of proposition 10

Proof. We can use the price approximation derived in 4.6.1.2 and derive the derivative of the convenience yield with respect to volatility. It is given by

$$\begin{aligned}
\frac{\partial \psi^i}{\partial \sigma^2} &= -\frac{\partial \psi^i}{\partial(1/\eta)}\frac{\gamma}{\frac{2}{1+\lambda}Y_1}\left(\widetilde{\tau} Q^f \frac{\partial P_1^f}{\partial \gamma\sigma^2} + \tau^m Q^m \frac{\partial P_1^m}{\partial \gamma\sigma^2}\right) \\
\frac{\partial \psi^i}{\partial \sigma^2} &= \frac{\partial \psi^i}{\partial(1/\eta)}\frac{\gamma}{\frac{2}{1+\lambda}Y_1}\left(\tau^m - \widetilde{\tau}\right) Q^f \frac{\partial P_1^f}{\partial \gamma\sigma^2}
\end{aligned}$$

where $\frac{\partial \psi^i}{\partial(1/\eta)} = -\left(\psi^i\right)^2 \frac{\tau^i \lambda(1-\lambda)}{(\lambda+(1-\lambda)/\eta)^2} < 0$, $\frac{\partial P_1^f}{\partial \gamma\sigma^2} = \frac{1}{Q^f}(1-\lambda)\frac{1}{1+\lambda}Y_1\frac{1}{\gamma\sigma^2}\left(\frac{\pi}{\gamma\sigma^2} - \frac{\partial \pi}{\partial \gamma\sigma^2}\right) > 0$, $\frac{\partial P_1^m}{\partial \gamma\sigma^2} = -\frac{\partial P_1^f}{\partial \gamma\sigma^2}\frac{Q^f}{Q^m} < 0$ and $\frac{\partial \pi}{\partial \gamma\sigma^2} = \frac{1}{2}\left(\mathcal{A}^2 + \frac{Q^m}{Q^f}\gamma\sigma^2\right)^{-\frac{1}{2}}\frac{Q^m}{Q^f} > 0$ for all $i \in \mathbb{I} \cup m$.

Note that $\frac{\pi}{\gamma\sigma^2} - \frac{\partial\pi}{\partial\gamma\sigma^2} > 0$. We can see this by rearranging the terms:

$$
\begin{aligned}
\frac{\pi}{\gamma\sigma^2} &> \frac{\partial\pi}{\partial\gamma\sigma^2} \\
\frac{1}{\gamma\sigma^2}\left(\mathcal{A} + \left(\mathcal{A}^2 + \frac{Q^2}{Q^1}\gamma\sigma^2\right)^{\frac{1}{2}}\right) &> \frac{1}{2}\left(\mathcal{A}^2 + \frac{Q^2}{Q^1}\gamma\sigma^2\right)^{-\frac{1}{2}}\frac{Q^2}{Q^1} \\
\frac{1}{\gamma\sigma^2}\frac{Q^1}{Q^2}\left(\mathcal{A}\sqrt{\mathcal{A}^2 + \frac{Q^2}{Q^1}\gamma\sigma^2} + \mathcal{A}^2\right) + 1 &> \frac{1}{2}.
\end{aligned}
$$

Analogously the derivative with respect to risk aversion is

$$
\frac{\partial\psi^i}{\partial\gamma} = \frac{\partial\psi^i}{\partial(1/\eta)}\frac{\sigma^2}{\frac{2}{1+\lambda}Y_1}\left(\tau^m - \widetilde{\tau}\right)Q^f\frac{\partial P_1^f}{\partial\gamma\sigma^2}
$$

for all $i \in \mathbb{I} \cup m$. □

Welfare

The social planner problem is given by

$$
\begin{aligned}
\max_{c_0, c_{1,\overline{S}}, c_{1,\underline{S}}} \; & log(c_0) + \lambda\left[2log(c_{1,\overline{S}})\right] + (1-\lambda)\left[log(c_{1,\underline{S}}) + log(Q^f + \varphi_2^m(z)Q^m)\right] \\
& \text{s.t.} \\
& Y_0 = c_0 \\
& Y_1 = \lambda c_{1,\overline{S}} + (1-\lambda)c_{1,\underline{S}}.
\end{aligned}
$$

The optimal consumption decisions are given by $c_0 = Y_0$, $c_{1,\overline{S}} = 2\frac{Y_1}{(1+\lambda)}$, and $c_{1,\underline{S}} = \frac{Y_1}{(1+\lambda)}$. We define $m_{1,\underline{S}} \equiv \frac{\frac{1}{c_{1,\underline{S}}}}{\frac{1}{c_0}}$ and $m_{1,\overline{S}} \equiv \frac{\frac{2}{c_{1,\overline{S}}}}{\frac{1}{c_0}}$. Solving the problem implies: $\eta = \frac{m_{1,\underline{S}}}{m_{1,\overline{S}}} = 1$.

The price of convenience can be used as a welfare statistic. To see this, define welfare as the objective function of the social planner and insert the constraints. This yields

$$
W \equiv log(Y_0) + 2\lambda log\left[\frac{1}{\lambda}\left(Y_1 - (1-\lambda)c_{1,\underline{S}}\right)\right] + (1-\lambda)\left[log(c_{1,\underline{S}}) + log(Q^f + \varphi_2^m(z)Q^m)\right].
$$

We can express $c_{1,\underline{S}}$ as a function of the price of convenience, $c_{1,\underline{S}} = \frac{Y_1}{2\lambda\eta + (1-\lambda)}$. We insert it. This implies:

$$
\begin{aligned}
W \equiv\; & log\left(Y_0\right) + 2\lambda log\left(\frac{1}{\lambda}\right) + 2\lambda log\left(Y_1 - (1-\lambda)\frac{Y_1}{2\lambda\eta + (1-\lambda)}\right) \\
& + (1-\lambda)log\left(\frac{Y_1}{2\lambda\eta + (1-\lambda)}\right) + (1-\lambda)log\left(Q^f + \varphi_2^m(z)Q^m\right).
\end{aligned}
$$

The first order condition is given by:

$$\frac{\partial W}{\partial \eta} = \left(\frac{1}{\eta}(1-\lambda) - (1-\lambda)\right)\frac{2\lambda}{2\lambda\eta + (1-\lambda)}.$$

It follows that $\frac{\partial W}{\partial \eta} = 0$ iff $\eta = 1$. We observe that for $\eta \geq 1$[25] welfare is uniformly decreasing in the price of convenience, i.e. $\frac{\partial W}{\partial \eta} < 0$. Therefore the price of convenience can be used as a welfare statistic with maximal welfare being reached if $\eta = 1$.

Proposition 11 (**Welfare**). *For $\eta \geq 1$ welfare W is uniformly decreasing in the price of convenience, η, $\frac{\partial W}{\partial \eta} < 0$, and maximal welfare is reached if $\eta = 1$.*

Price approximation

If $\left(\frac{P_1^f - P_1^m}{P_1^m}\right)$ is small, then we can find an approximative closed form solution for the prices. It is convenient to define $\mathcal{A} \equiv \frac{1}{2}\left[log\left[E\left(\frac{\varphi_2^m(z)}{\varphi_2^f}\right)\right] - 1 - \frac{Q^m}{Q^f}\right]$ for this. We state the following lemma:

Lemma 6 (**Prices**). *If $log\left(1 + \frac{P_1^f - P_1^m}{P_1^m}\right) \approx \left(\frac{P_1^f - P_1^m}{P_1^m}\right)$ holds, then the prices are given by $P_1^f = \frac{1}{Q^f}(1-\lambda)(1-\frac{\pi}{\gamma\sigma^2})\frac{1}{1+\lambda}Y_1$ and $P_1^m = \frac{1}{Q^m}(1-\lambda)\frac{\pi}{\gamma\sigma^2}\frac{1}{1+\lambda}Y_1$ with $\pi = \mathcal{A} + \sqrt{\mathcal{A}^2 + \frac{Q^m}{Q^f}\gamma\sigma^2}$.*

Proof. To prove this lemma we start with the perceived risk premium on the risky asset which is given by $\pi = E\left[log\left(\frac{\varphi_2^m(z)}{P_1^m}\right)\right] + \frac{\sigma^2}{2} - log\left(\frac{\varphi_2^f}{P_1^f}\right)$. We can rewrite it as follows given the assumption that $log\left(1 + \frac{P_1^f - P_1^m}{P_1^m}\right) \approx \left(\frac{P_1^f - P_1^m}{P_1^m}\right)$:

$$\begin{aligned}
\pi &= log\left[E\left(\frac{\varphi_2^m(z)}{\varphi_2^f}\frac{P_1^f}{P_1^m}\right)\right] \\
\pi &= log\left[E\left(\frac{\varphi_2^m(z)}{\varphi_2^f}\right)\right] + log\left[1 + \frac{P_1^f - P_1^m}{P_1^m}\right] \\
\pi &= log\left[E\left(\frac{\varphi_2^m(z)}{\varphi_2^f}\right)\right] + \frac{P_1^f - P_1^m}{P_1^m}.
\end{aligned}$$

Therefore $\pi = \Psi + \frac{P_1^f}{P_1^m}$ where $\Psi \equiv log\left[E\left(\frac{\varphi_2^m(z)}{\varphi_2^f}\right)\right] - 1$. Using the equilibrium price equations it follows that prices are given by

$$\begin{aligned}
P_1^m Q^m &= \left(\Psi + \frac{P_1^f}{P_1^m}\right)\Omega \\
P_1^f Q^f &= \left(\gamma\sigma^2 - \left(\Psi + \frac{P_1^f}{P_1^m}\right)\right)\Omega
\end{aligned}$$

[25] It is possible for the social planner to set $\eta < 1$ but in our model this is not a possible equilibrium outcome.

where $\Omega \equiv (1-\lambda)\frac{Y_1+A_1}{2}\frac{1}{\gamma\sigma^2}$. From summing both equations it follows that $P_1^f Q^f = \gamma\sigma^2\Omega - P_1^m Q^m$. We rearrange the second equation and insert for P_1^f:

$$\begin{aligned}
P_1^m Q^m &= \left(\Psi + \frac{P_1^f}{P_1^m}\right)\Omega \\
(P_1^m)^2 - P_1^m \Psi \frac{\Omega}{Q^m} &= P_1^f \frac{\Omega}{Q^m} \\
(P_1^m)^2 - P_1^m \Psi \frac{\Omega}{Q^m} &= \left(\frac{\gamma\sigma^2\Omega - P_1^m Q^m}{Q^f}\right)\frac{\Omega}{Q^m} \\
(P_1^m)^2 + P_1^m \left(\frac{1}{Q^f} - \frac{\Psi}{Q^m}\right)\Omega - \gamma\sigma^2 \frac{\Omega^2}{Q^f Q^m} &= 0.
\end{aligned}$$

The function $f(P_1^m) = (P_1^m)^2 + P_1^m\left(\frac{1}{Q^f} - \frac{\Psi}{Q^m}\right)\Omega - \gamma\sigma^2\frac{\Omega^2}{Q^f Q^m}$ has two roots. If $P_1^m = 0$ then $f(P_1^m) = -\gamma\sigma^2\frac{\Omega^2}{Q^f Q^m}$. Therefore at one root P_1^m must be positive and at the other negative. We are interested in the positive P_1^m. Therefore P_1^m is given by

$$\begin{aligned}
P_1^m &= \frac{-\left(\frac{1}{Q^f} - \frac{\Psi}{Q^m}\right)\Omega + \sqrt{\left(\frac{1}{Q^f} - \frac{\Psi}{Q^m}\right)^2\Omega^2 + 4\gamma\sigma^2\frac{\Omega^2}{Q^f Q^m}}}{2} \\
P_1^m &= \frac{\Omega}{2}\left(\frac{\Psi}{Q^m} - \frac{1}{Q^f}\right) + \Omega\sqrt{\frac{1}{4}\left(\frac{\Psi}{Q^m} - \frac{1}{Q^f}\right)^2 + \frac{\gamma\sigma^2}{Q^f Q^m}}.
\end{aligned}$$

We know that $\pi = \frac{P_1^m Q^m}{\Omega}$. Therefore $\pi = \frac{1}{2}\left(\Psi - \frac{Q^m}{Q^f}\right) + \sqrt{\frac{1}{4}\left(\Psi - \frac{Q^m}{Q^f}\right)^2 + \gamma\sigma^2\frac{Q^m}{Q^f}}$. The prices are given by

$$\begin{aligned}
P_1^f &= \frac{1}{Q^f}(1-\lambda)(1-\frac{\pi}{\gamma\sigma^2})\frac{1}{1+\lambda}Y_1 \\
P_1^m &= \frac{1}{Q^m}(1-\lambda)\frac{\pi}{\gamma\sigma^2}\frac{1}{1+\lambda}Y_1
\end{aligned}$$

with

$$\pi = \mathcal{A} + \sqrt{\mathcal{A}^2 + \frac{Q^m}{Q^f}\gamma\sigma^2}$$

where $\mathcal{A} \equiv \frac{1}{2}\left[log\left[E\left(\frac{\varphi_2^m(z)}{\varphi_2^f}\right)\right] - 1 - \frac{Q^m}{Q^f}\right]$. □

Transaction cost microfoundation

As we saw in section 4.4.2, long duration assets are less convenient than shorter duration assets. In this section, we provide a microfoundation for this observation. It also justifies our assumption that high transaction costs must be associated with long maturity assets in our model. We use our original model and make the following adjustment to explicitly account for short and long maturity assets: We assume (for simplicity) that there are only two risk-free assets with payoff 1 in the economy. One matures after one period and the other after two periods. We denote the short maturity asset by l ("low") and the long maturity asset by h ("high"). We will give two examples of why the long maturity asset is less convenient relative to the short maturity asset.

Example 1

First we assume that if an asset is sold before maturing, the agents incur a fee. As this can only happen for the long maturity asset, the fee is denoted by τ^h. We start by describing the optimal behaviour in period 1. As in our standard model, the liquidity constrained agent will liquidate all assets in period 1 and therefore $X^l_{2,\overline{S}} = X^h_{2,\overline{S}} = 0$. The other kind of agents solve $\max_{X^h_{2,\underline{S}}} log(Y_1 + A_1 - P^h_1 X^h_{2,\underline{S}}) + log(X^h_{2,\underline{S}})$, where $A_1 \equiv X^l_1 + P^h_1 X^h_1$. The first order condition is given by $\frac{P^h_1}{Y_1+A_1-P^h_1 X^h_{2,\underline{S}}} = \frac{1}{X^h_{2,\underline{S}}}$, or $P^h_1 X^h_{2,\underline{S}} = \frac{Y_1+A_1}{2}$. Market clearing implies $\lambda(X^h_{2,\overline{S}} - X^h_1) + (1-\lambda)(X^h_{2,\underline{S}} - X^h_1) = 0$, or $X^h_{2,\underline{S}} = \frac{X^h_1}{(1-\lambda)}$. In period 0 the agents solve the following maximisation problem:

$$\max_{c_0, X^l_1, X^h_1} log(c_0) + \lambda\chi V_{1,\overline{S}} + (1-\lambda)V_{1,\underline{S}}$$

$$\text{s.t.}$$

$$Y_0 = c_0 + P^l_0 X^l_1 + P^h_0 X^h_1$$

where $V_{1,\overline{S}} = log(Y_1 + A_1 - \tau^h P^h_1 X^h_1)$ and $V_{1,\underline{S}} = log\left(Y_1 + A_1 - \frac{P^h_1 X^h_1}{1-\lambda}\right) + log\left(\frac{X^h_1}{1-\lambda}\right)$, or $V_{1,\underline{S}} = log\left[Y_1 + A_1 - \frac{Y_1+A_1}{2}\right] + log\left[\frac{Y_1+A_1}{2}\right]$. The first order conditions are given by

$$\frac{1}{Y_0 - P^l_0 X^l_1 - P^h_0 X^h_1} = \lambda\left(\frac{\chi}{Y_1 + A_1 - \tau^h X^h_1}\frac{1}{P^l_0}\right) + (1-\lambda)\left(\frac{2}{Y_1 + A_1}\frac{1}{P^l_0}\right)$$

and

$$\frac{1}{Y_0 - P^l_0 X^l_1 - P^h_0 X^h_1} = \lambda\left(\frac{\chi}{Y_1 + A_1 - \tau^h X^h_1}\frac{P^h_1}{P^h_0}(1-\tau^h)\right) + (1-\lambda)\left(\frac{2}{Y_1 + A_1}\frac{P^h_1}{P^h_0}\right).$$

Market clearing in period 0 implies that $X^l_1 = Q^l$ and $X^h_1 = Q^h$. From combining the first order condition of the liquidity unconstrained agent in period 1 and market clearing it follows that $P^h_1 = \frac{1}{\frac{2}{(1-\lambda)}-1}\frac{Y_1+Q^l}{Q^h}$. From now on we assume that $\chi = 2$ (as in the main body). In addition we define $R^h_1 \equiv \frac{P^h_1}{P^h_0}$ and $R^l_1 \equiv \frac{1}{P^l_0}$. Note that $\frac{\partial V_{1,s}}{\partial A_1} = \frac{\partial U_{1,s}}{\partial c_{1,s}}$ and rewriting the first of the two first order conditions implies $\frac{1}{R^l_1} = E_s\left(\frac{\frac{\partial U_{1,s}}{\partial c_{1,s}}}{\frac{\partial U_1}{\partial c_0}}\right)$. As a next step we combine the first order conditions in period 1. To rearrange we use $E_s\left(\frac{\partial U_{1,s}}{\partial c_{1,s}}\right) \equiv \lambda\left(\frac{\partial U_{1,\overline{S}}}{\partial c_{1,\overline{S}}}\right) + (1-\lambda)\left(\frac{\partial U_{1,\underline{S}}}{\partial c_{1,\underline{S}}}\right)$. Rearranging yields

$$\begin{aligned}
0 &= \lambda \frac{2}{Y_1 + A_1}\left[R_1^l - R_1^h + R_1^h \tau^h\right] + (1-\lambda)\left[\frac{2}{Y_1 + A_1}(R_1^l - R_1^h)\right] \\
0 &= E_s\left\{\frac{\partial U_{1,s}}{\partial c_{1,s}}\left[1 - \frac{R_1^h}{R_1^l}(1 + \mathcal{I}_{s=\overline{S}}\tau^h)\right]\right\} \\
0 &= COV\left[\frac{\partial U_{1,s}}{\partial c_{1,s}}, \mathcal{I}_{s=\overline{S}}\tau^h\right]\frac{R_1^h}{R_1^l} + E_s\left[\frac{\partial U_{1,s}}{\partial c_{1,s}}\right]\left[1 - \frac{R_1^h}{R_1^l}(1 - \lambda\tau^h)\right] \\
0 &= COV\left[m_{1,s}, \mathcal{I}_{s=\overline{S}}\tau^h\right]\frac{R_1^h}{R_1^l} + E_s\left[m_{1,s}\right]\left[1 - \frac{R_1^h}{R_1^l}(1 - \lambda\tau^h)\right] \\
0 &= \frac{1}{R_1^l} + \left[COV\left[m_{1,s}, \mathcal{I}_{s=\overline{S}}\tau^h\right] - E_s\left[m_{1,s}\right](1 - \lambda\tau^h)\right]\frac{R_1^h}{R_1^l}.
\end{aligned}$$

Further rearranging yields

$$\frac{R_1^h}{R_1^l} = \frac{1}{R_1^l}\frac{1}{\left\{E_s\left[m_{1,s}\right](1-\lambda\tau^h) - COV\left[m_{1,s}, \mathcal{I}_{s=\overline{S}}\tau^h\right]\right\}}.$$

We know that $COV\left[m_{1,s}, \mathcal{I}_{s=\overline{S}}\tau^h\right] = \lambda\tau^h m_{\overline{S}} - E_s\left[m_{1,s}\right]\lambda\tau^h$. Therefore

$$\begin{aligned}
\frac{R_1^h}{R_1^l} &= \frac{1}{1 - \lambda\tau^h - \lambda\tau^h\left[m_{\overline{S}} - E_s\left[m_{1,s}\right]\right]R_1^l} > 0 \\
\frac{R_1^h}{R_1^l} &= \frac{1}{1 - \lambda\tau^h - \lambda\tau^h(1-\lambda)(\eta - 1)m_{\underline{S}}R_1^l} > 0
\end{aligned}$$

or rewritten

$$\frac{R_1^h}{R_1^l} = \frac{1}{1 - \lambda\tau^h - \lambda\tau^h(1-\lambda)(Y_0 - P_0^l Q_0^l - P_0^h Q_0^h)\left[\frac{2}{Y_1 + A_1 - \tau^h P_1^h Q^h} - \frac{2}{Y_1 + A_1}\right]R_1^l} > 0.$$

The long maturity asset is less convenient. The short maturity asset carries a convenience premium when being compared to the long maturity asset. The reason is that the long maturity asset has a lower probability to mature at the moment the liquidity is needed.

Example 2

Second, we assume that there is an aggregate uncertainty in $t = 1$. We assume that the endowment will be high or low in the first period. In period 0, the value of the endowment, denoted by Y_1^k, is unknown, but the agents know the distribution. We use this explicit example for illustrative purposes and later generalise the microfoundation. In the general version, the uncertainty can be about any exogenous variable.

Analogous to above and adjusted to the current example, the first order conditions are given by

$$\frac{1}{Y_0 - P_0^l X_1^l - P_0^h X_1^h} = E\left(\frac{2}{Y_1^k + A_1}\frac{1}{P_0^l}\right)$$

and

$$\frac{1}{Y_0 - P_0^l X_1^l - P_0^h X_1^h} = E\left(\frac{2}{Y_1^k + A_1}\frac{P_1^h}{P_0^h}\right).$$

We combine the first order conditions. Note that $\frac{\partial U_{1,\underline{S}}}{\partial A_1} = \frac{\partial U_{1,\overline{S}}}{\partial A_1}$ and $m_{1,\underline{S}} = m_{1,\overline{S}}$. We will use $\frac{1}{R_1^l} = E\left(\frac{\frac{\partial U_{1,\underline{S}}}{\partial A_1}}{\frac{\partial U_1}{\partial c_0}}\right)$. Therefore

$$\begin{aligned}
0 =& E\left[\frac{2}{Y_1^k + A_1}\left(R_1^l - R_1^h\right)\right] \\
0 =& E\left[\frac{\partial U_{1,\underline{S}}}{\partial c_{1,\underline{S}}}\left(1 - \frac{R_1^h}{R_1^l}\right)\right] \\
0 =& E\left[m_{1,\underline{S}}\left(1 - \frac{R_1^h}{R_1^l}\right)\right] \\
0 =& \frac{1}{R_1^l} - E\left(m_{1,\underline{S}}\frac{R_1^h}{R_1^l}\right) \\
0 =& \frac{1}{R_1^l} - COV\left(\frac{R_1^h}{R_1^l}, m_{1,\underline{S}}\right) - E\left(\frac{R_1^h}{R_1^l}\right)E\left(m_{1,\underline{S}}\right).
\end{aligned}$$

Further rearranging yields

$$\begin{aligned}
E\left(\frac{R_1^h}{R_1^l}\right) =& \frac{\frac{1}{R_1^l} - COV\left(m_{1,\underline{S}}, \frac{R_1^h}{R_1^l}\right)}{E\left(m_{1,\underline{S}}\right)} \\
E\left(\frac{R_1^h}{R_1^l}\right) =& 1 - COV\left(m_{1,\underline{S}}, R_1^h\right) \\
E\left(\frac{R_1^h}{R_1^l}\right) =& 1 - COV\left(\frac{m_{1,\overline{S}}}{\eta}, R_1^h\right) \\
E\left(\frac{R_1^h}{R_1^l}\right) =& 1 - 2\,COV\left(\frac{Y_0 - P_0^l Q_0^l - P_0^h Q_0^h}{Y_1^k + A_1}, R_1^h\right).
\end{aligned}$$

From $P_1^h = \frac{1}{\frac{2}{1-\lambda} - 1}\frac{Y_1^k + Q^l}{Q^h}$ it follows that $\frac{\partial P_1^h}{\partial Y_1^k} > 0$. Therefore $COV\left[\frac{Y_0 - P_0^l Q_0^l - P_0^h Q_0^h}{Y_1^k + A_1}, R_1^h\right] < 0$. It follows that

$$E\left(\frac{R_1^h}{R_1^l}\right) > 1.$$

The long maturity asset is on average less convenient because in period $t = 1$ the asset has a high price, or return when the marginal utility is low.

Generalisation

Finally, we can derive (following the same steps as in the examples above) a more general function that incorporates both channels, transaction costs and aggregate uncertainty, and further generalises the latter. First, we generalise by using an isoelastic utility function instead of a log utility function. Second, we allow agents to discount the second period with the factor $\beta \leq 1$. Third, we now leave open which variable x^i is affected by an aggregate uncertainty in period 0 that is revealed in period 1. For example, we could introduce the payoff risk as an aggregate risk or an uncertainty in the discount factor β. The aggregate uncertainty in the endowment was just an example.

The result that the covariance and not any volatility implies the effect of aggregate uncertainty on the convenience return is more general. We derive a general function for the convenience yields (which nests the convenience yields of the above examples):

$$E\left(\frac{R_1^h}{R_1^l}\right) = \frac{1 + COV\left[COV\left[m_{1,s}, \mathcal{I}_{s=\overline{S}}\tau^h\right] - E_s\left[m_{1,s}\right](1-\lambda\tau^h), R_1^h\right]}{1 - \lambda\tau^h - E\left\{COV\left[m_{1,s}, \mathcal{I}_{s=\overline{S}}\tau^h\right] R_1^l\right\}},$$

or

$$E\left(\frac{R_1^h}{R_1^l}\right) = \frac{1 + COV\left[\lambda\tau^h m_{\overline{S}} - E_s\left[m_{1,s}\right], R_1^h\right]}{1 - \lambda\tau^h - \lambda\tau^h E\left[m_{\overline{S}} - E_s\left[m_{1,s}\right]\right] R_1^l}$$

$$E\left(\frac{R_1^h}{R_1^l}\right) = \frac{1 + COV\left[\left[\lambda(\tau^h - 1)\eta - (1-\lambda)\right] m_{\underline{S}}, R_1^h\right]}{1 - \lambda\tau^h - \lambda\tau^h E\left[(1-\lambda)(\eta - 1)m_{\underline{S}}\right] R_1^l}.$$

4.6.2 Data

We obtain government security yields and OIS rates from Bloomberg (see table 4.5). We use it to calculate the Treasury-OIS spreads. The maturities we use are 3 months, 1 year, 2 years, 5 years, 7 years and 10 years. The frequency is daily, which we aggregate to weekly. As already mentioned, we also use primary dealer net positions. We obtain these from the public primary dealer statistics of the New York FED. We have data on Treasury bills and bonds. We divide the bond data into the following maturity baskets: (0y, 1y], (1y, 3y], (3y, 6y], and (6y, 11y]. The frequency is weekly. Data on total outstanding debt of the US government and outstanding bills are obtained from Bloomberg (tickers: PUBLDEBT Index and DEBPBILL Index). The frequency is daily for the former and monthly for the latter. The MOVE Index also comes from Bloomberg (ticker: MOVE Index), and the VIX and the SP 500 Index are obtained from FRED. The frequency is daily, which we aggregate to weekly.[26] Lastly, we obtain the Effective

[26] When we change the frequency of the data, we always take the last available day in the new time unit for the new frequency. The only exception is when we match the convenience yields to the primary dealer net positions. There we match the convenience yields, the SP 500 Index and the Effective Federal Funds Rate to the same date

Federal Funds Rate from FRED. The frequency is daily, which we aggregate to weekly. This following table gives an overview over the Bloomberg tickers of the used data series for the Treasury yields and OIS rates.

Table 4.5: Data sources

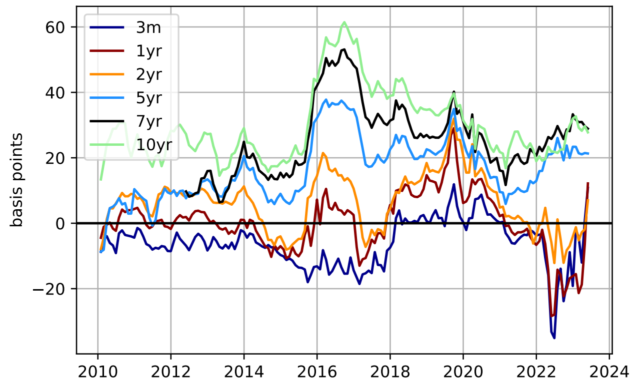
Data	Maturity	Source	Ticker
Treasury	3M	Bloomberg	USGG3M Index
	1Y	Bloomberg	USGG1Y Index
	2Y	Bloomberg	USGG2Y Index
	5Y	Bloomberg	USGG5Y Index
	7Y	Bloomberg	USGG7Y Index
	10Y	Bloomberg	USGG10Y Index
OIS	3M	Bloomberg	USSOC Curncy
	1Y	Bloomberg	USSO1 Curncy
	2Y	Bloomberg	USSO2 Curncy
	5Y	Bloomberg	USSO5 Curncy
	7Y	Bloomberg	USSO7 Curncy
	10Y	Bloomberg	USOSFR10 Curncy

as the net positions. When we adjust the frequency of the time series plotted in the figures, we use the average instead of the last available day, as this gives a more accurate overview and is uncritical to do, as no different time series are matched. The only exception is the figure 4.5, where we use the last day because we are comparing two series.

4.6.3 Omitted figures

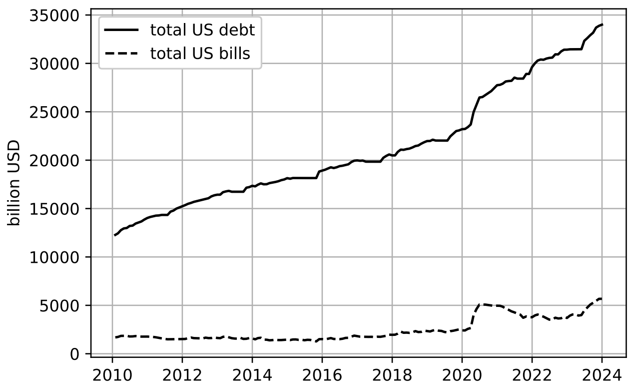
This section plots several omitted figures to which we refer in the main body.

Figure 4.4: Treasury-OIS spread



Notes: This figure shows the US Treasury-OIS spread across different maturities. Source: Bloomberg and authors calculations.

Figure 4.5: Total outstanding US government debt and bills



Notes: This figure shows the monthly total outstanding US government debt and total oustanding US government bills. Source: Bloomberg and authors calculations.

4.6.4 Omitted tables

This section plots several omitted tables to which we refer in the main body.

Variation in supply: primary dealer net positions - individual maturity regressions

Table 4.6: Impact of supply fluctuations (measured by net positions) on the convenience yields of 1 year Treasury bonds

	(1)	(2)
	TO_{1Y}-TO_{3M}	TO_{1Y}-TO_{3M}
Positions with $\tau < \tilde{\tau}$	-0.2379***	-0.1109
	(0.120)	(8.663)
Positions with $\tau > \tilde{\tau}$	0.8555***	1.0396***
	(0.174)	(8.663)
ln(SP 500)		-5.5316***
		(1.131)
Effective Federal Funds Rate		1.4770***
		(0.738)
Constant	7.2937***	47.9268***
	(0.857)	(8.663)
N	626	617
R^2	0.123	0.235

Notes: The outcome variable consists of the convenience yields of 1 year Treasury bonds, denoted by $TO_{1Y,t} - TO_{3M,t}$. The explanatory variables are the deflated primary dealers' net positions in Treasuries above and below the volume weighted average maturity separately, the SP 500 Index and the Effective Federal Funds Rate. The frequency is weekly and the time horizon is 2010-2021. Standard errors are heteroscedasticity and autocorrelation robust (HAC) using 6 lags and without small sample correction. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Table 4.7: Impact of supply fluctuations (measured by net positions) on the convenience yields of 2 year Treasury bonds

	(1)	(2)
	TO_{2Y}-TO_{3M}	TO_{2Y}-TO_{3M}
Positions with $\tau < \tilde{\tau}$	-0.1223	0.0775
	(0.209)	(0.200)
Positions with $\tau > \tilde{\tau}$	0.7376***	1.5858***
	(0.230)	(0.463)
ln(SP 500)		-9.6972***
		(1.880)
Effective Federal Funds Rate		-0.0042
		(1.071)
Constant	12.5176***	84.5346***
	(1.418)	(14.694)
N	626	617
R^2	0.046	0.185

Notes: The outcome variable consists of the convenience yields of 2 year Treasury bonds, denoted by $TO_{2Y,t} - TO_{3M,t}$. The explanatory variables are the deflated primary dealers' net positions in Treasuries above and below the volume weighted average maturity separately, the SP 500 Index and the Effective Federal Funds Rate. The frequency is weekly and the time horizon is 2010-2021. Standard errors are heteroscedasticity and autocorrelation robust (HAC) using 6 lags and without small sample correction. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Table 4.8: Impact of supply fluctuations (measured by net positions) on the convenience yields of 5 year Treasury bonds

	(1)	(2)
	TO_{5Y}-TO_{3M}	TO_{5Y}-TO_{3M}
Positions with $\tau < \tilde{\tau}$	-1.0791***	-1.1245***
	(0.358)	(0.324)
Positions with $\tau > \tilde{\tau}$	2.0051***	1.9301***
	(0.352)	(0.735)
ln(SP 500)		2.3997
		(3.619)
Effective Federal Funds Rate		-0.5407
		(1.673)
Constant	23.9821***	6.2322
	(2.401)	(28.222)
N	623	614
R^2	0.173	0.179

Notes: The outcome variable consists of the convenience yields of 5 year Treasury bonds, denoted by $TO_{5Y,t} - TO_{3M,t}$. The explanatory variables are the deflated primary dealers' net positions in Treasuries above and below the volume weighted average maturity separately, the SP 500 Index and the Effective Federal Funds Rate. The frequency is weekly and the time horizon is 2010-2021. Standard errors are heteroscedasticity and autocorrelation robust (HAC) using 6 lags and without small sample correction. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Table 4.9: Impact of supply fluctuations (measured by net positions) on the convenience yields of 7 year Treasury bonds

	(1)	(2)
	TO_{7Y}-TO_{3M}	TO_{7Y}-TO_{3M}
Positions with $\tau < \tilde{\tau}$	-2.7912***	-2.8613***
	(0.347)	(0.334)
Positions with $\tau > \tilde{\tau}$	2.7304***	5.0495***
	(0.499)	(0.850)
ln(SP 500)		-11.9130***
		(3.521)
Effective Federal Funds Rate		-5.0498***
		(1.683)
Constant	41.4807***	133.5287***
	(2.779)	(27.817)
N	503	494
R^2	0.412	0.475

Notes: The outcome variable consists of the convenience yields of 7 year Treasury bonds, denoted by $TO_{7Y,t} - TO_{3M,t}$. The explanatory variables are the deflated primary dealers' net positions in Treasuries above and below the volume weighted average maturity separately, the SP 500 Index and the Effective Federal Funds Rate. The frequency is weekly and the time horizon is 2010-2021. Standard errors are heteroscedasticity and autocorrelation robust (HAC) using 6 lags and without small sample correction. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Table 4.10: Impact of supply fluctuations (measured by net positions) on the convenience yields of 10 year Treasury bonds

	(1)	(2)
	TO_{10Y}-TO_{3M}	TO_{10Y}-TO_{3M}
Positions with $\tau < \tilde{\tau}$	-1.8880***	-1.8429***
	(0.370)	(0.360)
Positions with $\tau > \tilde{\tau}$	2.2344***	2.5309***
	(0.422)	(0.767)
ln(SP 500)		-1.7357
		(3.487)
Effective Federal Funds Rate		0.6235
		(1.786)
Constant	42.2483***	55.2768***
	(2.522)	(26.886)
N	626	617
R^2	0.247	0.248

Notes: The outcome variable consists of the convenience yields of 10 year Treasury bonds, denoted by $TO_{10Y,t} - TO_{3M,t}$. The explanatory variables are the deflated primary dealers' net positions in Treasuries above and below the volume weighted average maturity separately, the SP 500 Index and the Effective Federal Funds Rate. The frequency is weekly and the time horizon is 2010-2021. Standard errors are heteroscedasticity and autocorrelation robust (HAC) using 6 lags and without small sample correction. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Variation in supply: debt ceiling dates - individual maturity regressions

Table 4.11: Impact of supply fluctuations (measured by debt ceilings) on the convenience yields of 1 year Treasury bonds

	Debt growth	TO_{1Y}-TO_{3M}
Debt ceiling dates	-0.1726***	
	(0.018)	
Debt growth		12.853***
		(4.1419)
Constant	0.1756***	4.9127***
	(0.018)	(0.6487)
Model	First-stage	IV-2SLS
N	697	697
F-Statistic	93.64	

Notes: The outcome variable consists of the convenience yields of 1 year Treasury bonds, denoted by $TO_{1Y,t} - TO_{3M,t}$. The explanatory variable consists of the fitted values of debt growth. The instrument are all days on which daily debt growth was constant for at least five weeks. The frequency is weekly and the time horizon is 2010-2023. Standard errors are heteroscedasticity robust. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Table 4.12: Impact of supply fluctuations (measured by debt ceilings) on the convenience yields of 2 year Treasury bonds

	Debt growth	TO_{2Y}-TO_{3M}
Debt ceiling dates	-0.1726***	
	(0.018)	
Debt growth		20.508***
		(4.5336)
Constant	0.1756***	10.077***
	(0.018)	(0.6385)
Model	First-stage	IV-2SLS
N	697	697
F-Statistic	93.64	

Notes: The outcome variable consists of the convenience yields of 2 year Treasury bonds, denoted by $TO_{2Y,t} - TO_{3M,t}$. The explanatory variable consists of the fitted values of debt growth. The instrument are all days on which daily debt growth was constant for at least five weeks. The frequency is weekly and the time horizon is 2010-2023. Standard errors are heteroscedasticity robust. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Table 4.13: Impact of supply fluctuations (measured by debt ceilings) on the convenience yields of 5 year Treasury bonds

	Debt growth	TO_{5Y}-TO_{3M}
Debt ceiling dates	-0.1726***	
	(0.018)	
Debt growth		0.5971
		(5.3694)
Constant	0.1756***	22.933***
	(0.018)	(0.7086)
Model	First-stage	IV-2SLS
N	697	697
F-Statistic	93.64	

Notes: The outcome variable consists of the convenience yields of 5 year Treasury bonds, denoted by $TO_{5Y,t} - TO_{3M,t}$. The explanatory variable consists of the fitted values of debt growth. The instrument are all days on which daily debt growth was constant for at least five weeks. The frequency is weekly and the time horizon is 2010-2023. Standard errors are heteroscedasticity robust. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Table 4.14: Impact of supply fluctuations (measured by debt ceilings) on the convenience yields of 7 year Treasury bonds

	Debt growth	TO_{7Y}-TO_{3M}
Debt ceiling dates	-0.1624***	
	(0.021)	
Debt growth		9.2063
		(7.3568)
Constant	0.1640***	31.437***
	(0.021)	(0.8641)
Model	First-stage	IV-2SLS
N	576	576
F-Statistic	60.24	

Notes: The outcome variable consists of the convenience yields of 7 year Treasury bonds, denoted by $TO_{7Y,t} - TO_{3M,t}$. The explanatory variable consists of the fitted values of debt growth. The instrument are all days on which daily debt growth was constant for at least five weeks. The frequency is weekly and the time horizon is 2010-2023. Standard errors are heteroscedasticity robust. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Table 4.15: Impact of supply fluctuations (measured by debt ceilings) on the convenience yields of 10 year Treasury bonds

	Debt growth	TO_{10Y}-TO_{3M}
Debt ceiling dates	-0.1726***	
	(0.018)	
Debt growth		2.7987
		(6.5052)
Constant	0.1756***	35.926***
	(0.018)	(0.9255)
Model	First-stage	IV-2SLS
N	697	697
F-Statistic	0.072	

Notes: The outcome variable consists of the convenience yields of 10 year Treasury bonds, denoted by $TO_{10Y,t} - TO_{3M,t}$. The explanatory variable consists of the fitted values of debt growth. The instrument are all days on which daily debt growth was constant for at least five weeks. The frequency is weekly and the time horizon is 2010-2023. Standard errors are heteroscedasticity robust. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Effects of transaction costs and risk - individual maturity regressions

Table 4.16: Impact of the MOVE Index and the VIX on the convenience yields of 1 year Treasury bonds

	(1)	(3)
	TO_{1Y}-TO_{3M}	TO_{1Y}-TO_{3M}
ln(MOVE)	-0.6099	-2.9855**
	(1.618)	(1.768)
ln(VIX)	-1.3549	0.1772
	(1.191)	(1.107)
ln(SP 500)		-4.4953***
		(0.875)
Effective Federal Funds Rate		2.1842***
		(0.754)
Constant	13.5444***	52.5072***
	(6.208)	(12.158)
N	677	677
R^2	0.008	0.108

Notes: The outcome variable consists of the convenience yields of 1 year Treasury bonds, denoted by $TO_{1Y,t} - TO_{3M,t}$. The explanatory variables are the MOVE Index and the VIX, the SP 500 Index and the Effective Federal Funds Rate. The frequency is weekly and the time horizon is 2010-2022. Standard errors are heteroscedasticity and autocorrelation robust (HAC) using 6 lags and without small sample correction. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Table 4.17: Impact of the MOVE Index and the VIX on the convenience yields of 2 year Treasury bonds

	(1)	(3)
	TO_{2Y}-TO_{3M}	TO_{2Y}-TO_{3M}
ln(MOVE)	6.0071***	4.0463*
	(2.050)	(2.490)
ln(VIX)	-2.3908*	-1.0235
	(1.633)	(1.641)
ln(SP 500)		-3.5811***
		(1.823)
Effective Federal Funds Rate		2.2277***
		(0.856)
Constant	-5.6208	24.9806
	(8.207)	(20.528)
N	677	677
R^2	0.034	0.082

Notes: The outcome variable consists of the convenience yields of 2 year Treasury bonds, denoted by $TO_{2Y,t} - TO_{3M,t}$. The explanatory variables are the MOVE Index and the VIX, the SP 500 Index and the Effective Federal Funds Rate. The frequency is weekly and the time horizon is 2010-2022. Standard errors are heteroscedasticity and autocorrelation robust (HAC) using 6 lags and without small sample correction. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Table 4.18: Impact of the MOVE Index and the VIX on the convenience yields of 5 year Treasury bonds

	(1)	(3)
	TO_{5Y}-TO_{3M}	TO_{5Y}-TO_{3M}
ln(MOVE)	14.1760***	18.6050***
	(4.183)	(29.954)
ln(VIX)	-12.1708***	-13.5016***
	(2.625)	(2.527)
ln(SP 500)		10.2988***
		(2.813)
Effective Federal Funds Rate		2.2353***
		(1.033)
Constant	-2.7046	-98.4708***
	(18.156)	(29.954)
N	677	677
R^2	0.102	0.260

Notes: The outcome variable consists of the convenience yields of 5 year Treasury bonds, denoted by $TO_{5Y,t} - TO_{3M,t}$. The explanatory variables are the MOVE Index and the VIX, the SP 500 Index and the Effective Federal Funds Rate. The frequency is weekly and the time horizon is 2010-2022. Standard errors are heteroscedasticity and autocorrelation robust (HAC) using 6 lags and without small sample correction. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Table 4.19: Impact of the MOVE Index and the VIX on the convenience yields of 7 year Treasury bonds

	(1)	(3)
	TO_{7Y}-TO_{3M}	TO_{7Y}-TO_{3M}
ln(MOVE)	24.4564***	24.0397***
	(4.198)	(4.076)
ln(VIX)	-12.8660***	-14.6378***
	(3.539)	(3.671)
ln(SP 500)		5.3507***
		(3.892)
Effective Federal Funds Rate		1.8353
		(1.067)
Constant	-34.1769***	-70.70**
	(17.417)	(32.281)
N	555	555
R^2	0.174	0.206

Notes: The outcome variable consists of the convenience yields of 7 year Treasury bonds, denoted by $TO_{7Y,t} - TO_{3M,t}$. The explanatory variables are the MOVE Index and the VIX, the SP 500 Index and the Effective Federal Funds Rate. The frequency is weekly and the time horizon is 2010-2022. Standard errors are heteroscedasticity and autocorrelation robust (HAC) using 6 lags and without small sample correction. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Table 4.20: Impact of the MOVE Index and the VIX on the convenience yields of 10 year Treasury bonds

	(1)	(3)
	TO_{10Y}-TO_{3M}	TO_{10Y}-TO_{3M}
ln(MOVE)	12.4264***	13.2441***
	(3.557)	(3.650)
ln(VIX)	-15.5335***	-15.2574***
	(3.469)	(3.544)
ln(SP 500)		2.5573
		(2.311)
Effective Federal Funds Rate		2.5697***
		(1.111)
Constant	27.8291**	2.2400
	(14.676)	(26.149)
N	0.110	0.147
R^2	677	677

Notes: The outcome variable consists of the convenience yields of 10 year Treasury bonds, denoted by $10\mathrm{Y}_{1,t} - \mathrm{TO}_{3M,t}$. The explanatory variables are the MOVE Index and the VIX, the SP 500 Index and the Effective Federal Funds Rate. The frequency is weekly and the time horizon is 2010-2022. Standard errors are heteroscedasticity and autocorrelation robust (HAC) using 6 lags and without small sample correction. *** indicates significance at the 5% level. ** indicates significance at the 10% level. * indicates significance at the 15% level.

Third Chapter

Credit and Anonymity

Credit and Anonymity[1]

co-authored with

Remo Taudien[2]

Abstract

It is commonly believed that borrowers cannot be anonymous in unsecured credit relations because anonymity heavily reduces the scope for punishment and therefore makes credit unfeasible except for very special circumstances. However, we demonstrate that credit is generally feasible even if borrowers are anonymous. In particular, we construct equilibria where borrowers use potentially multiple pseudonyms (such as usernames or wallet addresses) to interact with lenders. We assume that the complete history of past actions committed by a pseudonym is public but not the identity behind that pseudonym. While borrowers cannot be directly punished due to their anonymity, there is still scope for punishment. One possibility is based on the loss of reputation accumulated by a pseudonym over time. Another involves charging a fee to create pseudonyms. Although credit and anonymity are not mutually exclusive, we also show that maintaining a borrower's anonymity is costly.

[1]We thank Lukas Altermatt, Fernando Alvarez, Pierpaolo Benigno, Martin Brown, Harris Dellas, Lorenz Driussi, Janet Jiang, Charlie Kahn, Ricardo Lagos, Sebastian Merkel, Cyril Monnet, Dirk Niepelt, Remo Nyffenegger, Martina Pons, Randy Wright, and the participants of the Macroeconomics Reading Group Bern, the Reading Group of the Economic Theory Group of the University of Basel, the Brown Bag Seminar Bern, the Young Economist Conference, the Rice-LEMMA Monetary Conference, the 3rd Annual CBER Conference, and the Summer Workshop on Money, Banking, Payments, and Finance for helpful comments and discussions.

[2]University of Bern and Study Center Gerzensee

5.1 Introduction

There is general agreement that a) credit plays a crucial role in modern financial and monetary systems, and b) there is a strong desire for anonymity among its users. However, if we want to facilitate a credit contract between two parties, it is also widely believed that anonymity cannot be maintained. The idea is that absent any collateral, the borrowing party has to reveal its identity which allows the lender party to punish the borrower in case of default. This in turn provides the necessary incentives for repayment on the borrower side. The absence of anonymity in uncollaterlised credit is problematic for many different applications, in particular for the development of decentralised finance (DeFi[1]) credit applications, as one of the aspiring goals of DeFi is to maintain the anonymity of it's users. In this paper we demonstrate that credit and anonymity are, contrary to popular belief, compatible.

At the heart of the matter lies the question of what qualifies as anonymity. We distinguish two concepts of anonymity that we refer to as *strict anonymity* and *pseudonymity*. Under the former, a user is deemed anonymous if and only if there is no public record on actions committed by an agent. For instance, a cash economy is considered strictly anonymous as there is generally no public record of cash transactions. In contrast, a user is pseudonymous if and only if the identity of agents responsible for certain actions is unknown even if the entire history of actions of those agents is publicly known. The scenario we have in mind is one in which agents use *accounts* (or pseudonyms) when transacting with each other such as a wallet address.[2] These accounts are then used to negotiate and record credit contracts. The history of accounts is public information but it is private information to whom the accounts belong. Only the account owner knows which ones are his accounts. Practically, many blockchain related applications are pseudonymous, for example, lending and borrowing via smart contracts on Ethereum-based protocols like Compound, Aave, and Uniswap.

Except for some special cases (see literature review), credit and strict anonymity are not compatible as the lack of a public record (in particular of defaults) makes the punishment of defaulting borrowers impossible. Whether or not credit is feasible under pseudonymity is still largely unexplored.[3] The lack of research on this question is surprising since many areas of the internet operate under a pseudonymous regime where users interact with each other through

[1]A comprehensive overview of DeFi and how it compares to traditional (centralised) finance is provided in Qin et al. (2021).

[2]Other examples are email accounts, user names or gamer tags just to name a few.

[3]We focus on economic feasibility, not technical feasibility. For credit to attain technical feasibility under pseudonymity, it requires the technology to a) record all debt obligations between accounts, b) validate the fulfillment of these obligations, and c) establish a transparent and easily accessible public record of these obligations. With the rapid advancements in blockchain technology, the prospect of achieving this feasibility within the upcoming years appears promising, despite some challenges.

website accounts, blockchain wallets, or virtual avatars, while a large part of the user's activity is recorded and observable. We will therefore focus our attention on the economic feasibility and efficiency of pseudonymous credit systems. Our approach represents a middle ground between the two extremes often considered in the literature: strict anonymity (which excludes credit) and full information (where anonymity is absent).

Our model, which is populated by two types of agents called borrowers and lenders, has a simple structure. Each period is divided into two subperiods: the borrowing and the repayment stage. During the first, borrowers and lenders meet bilaterally and it is assumed that lenders can produce a particular good, the credit good, of which the borrowers derive instantaneous utility from consuming it. The key challenge arises from the fact that borrowers cannot produce anything of value for the lender on the spot. Instead, the borrower can only promise to repay the lender in the second stage, the repayment stage, by producing a different good, the settlement good, to reimburse the lender. This structure captures in a very simple way the main economic challenges in credit relations which is the lack of intertemporal commitment from the borrower's side.[4]

Different to standard models of credit, borrowers and lenders interact with each other using accounts (or pseudonyms). More precisely, borrowers and lenders have access to a record-keeping technology that perfectly tracks debt relations and repayment histories across different accounts, which are owned by borrowers. Borrowers can always create new accounts at zero cost and choose each period on which account to record actions. This record-keeping technology aligns with the previously mentioned notion of pseudonymity. Finally, to make the problem interesting, we assume that borrowers exit the economy stochastically and are replaced by new borrowers. Thus, at any given time, an account lacking any history could indicate either a "young" borrower, opening an account for the first time, or an "older" borrower who has opened a new account. The key challenge of establishing a credit system in such an environment is: How to punish borrowers who have defaulted on their debt if they can always "clear their history" by creating and using new accounts?

Throughout the paper we consider two cases: the first where accounts are costless to create and the second where agents need to pay a fee to open a new account.

Our main finding is that there always exists an equilibrium where credit is feasible even if agents are pseudoynmous and accounts are costless to create. Intuitively, those equilibria work in the following way: accounts, rather than borrowers, earn *reputation*. Reputation is a mapping

[4] What makes this formulation so attractive is the fact that the lenders in our model perform two roles simultaneously which, in reality, are usually performed by two different agents. To take the example of mortgages, the bank usually lends funds to a borrower who uses them to buy a house. In our model, the lender lends the goods directly and performs in this sense both tasks at the same time.

from the account's history. Consistently repaying debt increases reputation. A higher reputation implies a higher credit limit and therefore a higher level of feasible consumption. Accounts that have a history of defaulting on their debts are barred from borrowing in the future. Even though borrowers can always create new accounts and borrow again after defaulting, this is costly to do because newly created accounts have no reputation and can therefore borrow only little. In these equilibria, borrowers never endogenously default on their debt because the *value of reputation*, i.e. the difference in continuation values between an account with a given level of reputation compared to an account with no reputation, is sufficiently high. An important, and somewhat surprising, implication is that borrowers optimally use only one account despite having the option to use a second account in parallel with the first.[5] The rationale behind this is that using a second account incurs an opportunity cost of forfeiting the chance to use the primary account which allows both, to gain more reputation on that account and to consume more. The underlying assumption generating this effect is that matching with lenders is time-consuming.[6]

We also show what is required to construct such equilibria. First, accounts that have defaulted have to be punished by reducing the amount they can borrow in the future (in our case, we assume that no amount can ever be borrowed again). Second, lenders need to reject any off-equilibrium offers. The reason is that absent such a punishment, borrowers and lenders bilaterally agree to exchange the highest amount of credit such that the borrower is indifferent between repaying and defaulting. However, we show that those "not-too-tight" debt limits, using the language of Alvarez and Jermann (2000), cannot be part of an equilibrium. The reason is the following: when bargaining borrowers and lenders do not internalise how their choices affect the value of reputation in equilibrium. We then show that if debt limits are not-too-tight, then the value of reputation collapses to zero and therefore also debt limits. Therefore, by punishing deviation from the equilibrium amount, we can enlarge the set of incentive-feasible equilibria and construct equilibria where the value of reputation and debt limits are positive.[7] This suggests that lenders need to "artificially" lower the amount they lend so that, in equilibrium, reputation is sufficiently valuable to provide the incentives for borrowers to repay.

While our model shows that credit is always economically feasible, we show that maintaining credit in a pseudonymous environment is costly. This is due to a trade-off between the consumption of "older" borrowers (those with a lot of reputation) and of "younger" borrowers (those with little or no reputation). Intuitively, supporting a large volume of trade requires

[5]The fact that borrowers can always create new accounts if another account is flagged as a defaulter still has an important effect on the equilibrium as it lowers the cost of defaulting.

[6]Two accounts held by a borrower cannot be used to accumulate reputation by pretending to trade with each other. To gain reputation goods have to be produced and exchanged. Only lenders can produce these goods (this assumption ultimately introduces gains from trade into the model).

[7]See Bethune et al. (2018) for a detailed discussion.

a high value of reputation to prohibit borrowers from defaulting on their debts. But for reputation to be highly valued, it must be that consumption is restricted for those agents with little reputation, such as "young" borrowers. Generally, pseudonymity is costly because it is impossible to differentiate between "young" borrowers and those who have defaulted, and, as a result, any punishment scheme affects borrowers not only off-equilibrium (if they default) but also on-equilibrium (when they are "young").

Finally, we discuss the case where agents need to pay a fee in order to open an account. The idea is that there is an authority that manages those accounts and charges a fee whenever an agent wants to open an account. We differentiate between two subcases: a first where those fees are a real cost and a second where the fees consist of collected goods and are redistributed. We show that the incentives for repayment are now directly related to those fees. To be precise, the debt limit is now a linear function of those fees and in the second case, also of the transfers the borrowers receive. In equilibrium, borrowers do not default because this would mean that they would have to pay the fee for a new account and, in the second case, they also forgo the transfer. We show that in both subcases we can always find equilibria for some traded quantities that exist. In a second step we analyse the optimal equilibrium for the case when the authority can choose the fee and redistribute it. We find that the quantity traded in the optimal costly accounts equilibrium is always higher than in the optimal reputation equilibrium in all reputation stages.

This second approach with costly accounts provides a high technological feasibility while the case with costless accounts speaks to the idea of making accounts on platforms accessible for everyone (independent of their budget).

Literature review Our model builds on the long tradition of search models. In particular, our model is based on Lagos and Wright (2005) which has been used for many different applications, such as monetary economics, banking and finance.[8] Examples of models of credit using that framework are Gu et al. (2013), Lotz and Zhang (2015), Carapella and Williamson (2015) and Gu et al. (2016) just to name a few. A feature of many of these models is that credit is not feasible under anonymity. As mentioned above, we innovate by using a different notion of anonymity.

A notable exception is Araujo (2004) in which he shows that a credit equilibrium is still feasible even with strict anonymity. Credit in his model is maintained by a trigger-strategy propagated through "word-of-mouth". However, this only works if agents are sufficiently patient and the population is sufficiently small. Our equilibrium exists for infinitely many agents with

[8]A comprehensive review of this literature can be found in Lagos et al. (2017).

arbitrary discount factors.

The paper closest to ours is Wang and Li (2023) who also study credit equilibria in a weakly anonymous environment. We share some similar results but, and importantly, there are also some interesting and significant differences leading to complementary insights. Firstly, our environment is different. For example, borrowers only sometimes match with lenders in our model. This is important as it rules out some of the mechanisms used in their paper to prevent the use of multiple accounts. We present a mechanism that is robust in this respect.[9] Furthermore, we construct a different equilibrium in the sense that ours is a finite reputation equilibrium while theirs is an infinite reputation equilibrium.[10] This allows us to use backwards induction and derive more analytical results. For example, we show that such an equilibrium always exists and that these equilibria cannot be not-too-tight.

Friedman and Resnick (2004) discuss pseudonyms in a game theoretic context. They study an infinitely repeated prisoner's dilemma where, similar to our model, agents are pseudonymous. They conclude, as we do, that maintaining cooperation is costly as building up reputation is costly. Nevertheless, they miss many of the specifics of credit in their treatment that we believe are important to highlight.

The work of Kehoe and Levine (1993) and Alvarez and Jermann (2000) share with our work the feature that endogenous debt limits arise in equilibrium. Agents lack commitment, and they can thus only credibly promise to repay a certain amount of debt.

Our construction of reputation systems and the costs they bear is also reminiscent of the "starting small" literature. For example in Watson (1999) agents play repeatedly a game similar to a prisoner's dilemma where they cooperate or betray each other. Agents decide before the first game how payoffs evolve over time. There are high and low type agents with incomplete information about each other. He shows that cooperation between high types can be maintained regardless of initial beliefs about each other, as long as the relationship starts small enough in terms of what is at stake in the relationship. Hua and Watson (2022) study a similar game. Each period the first of two players chooses a so called trust level which determines what is at stake. Player two says if he wants to betray or corporate. They also show that in equilibrium, due to trade-offs, relationships start small in terms of payoffs and then increase in level until a maximum. This notion of "starting small" in both examples is similar to our set-up.

Another strand of literature studies the impact of credit information sharing and reporting. Elul and Gottardi (2015) study the effects and welfare impact of information restrictions on

[9]A key question in both papers is how to prevent borrowers to use secondary accounts. The approaches taken are altogether different.

[10]The terminology in Wang and Li (2023) is different. For example, their "increasing-credit-limit schemes depending on account age" is what we call reputation.

credit equilibria. Similar to our set-up, an entrepreneur's reputation is determined by his credit record on repayments and defaults. Brown and Zehnder (2007) demonstrate that information sharing in credit markets has especially strong effects on repayment in the absence of third-party enforcement and repeated trading relationships. Lastly, our work is related to Kocherlakota and Wallace (1998) who do not focus on pseudonymous record keeping, but on another form of limited record keeping, namely delayed record keeping.

The paper is structured as follows: Section 5.2 sets out the general environment, preferences and technology. Section 5.3 describes the equilibrium recursively and proves the existence of anonymous credit. Section 5.4 analyses an alternative approach with costly accounts.

5.2 The environment

Time is discrete and indexed by $t = 0, 1, \dots \infty$. There are two types of agents called *borrowers* and *lenders*. There is a measure one of each type.

Each period consists of two subperiods. During the first subperiod, agents enter the *borrowing stage* (BS) where each borrower matches bilaterally with a random lender with probability $\sigma \in (0, 1]$. Lenders can produce and sell a *credit good*, $q_t \in \mathbb{R}_0^+$, which the borrowers desire to consume.[11] At the end of the BS borrowers exit the economy stochastically with probability p and are immediately replaced by new borrowers entering the economy so that there is always a mass one of borrowers. In the second subperiod, agents enter the *repayment stage* (RS), in which agents stay matched until the end of the period. In this stage, borrowers can produce the *settlement good*, $y_t \in \mathbb{R}$, at a linear cost.[12] Lenders on the other hand receive linear utility from consuming the settlement good. Both, the credit and settlement goods are non-storable across subperiods. The borrower's and lender's expected lifetime utility are given by

$$U_t^b = \sum_{j=0}^{\infty} \beta^j \mathbb{E}_t[u(q_{t+j}) - y_{t+j}],$$

$$U_t^l = \sum_{j=0}^{\infty} \beta^j \mathbb{E}_t[-c(q_{t+j}) + y_{t+j}],$$

where $\beta \in (0, 1)$ is the *discount factor*. Moreover, $u(q)$ is $\mathcal{C}^2$ and satisfies $u'(q) > 0$, $u''(q) < 0$, $u(0) = 0$, $\lim_{q\to\infty} u'(q) = 0$ and $\lim_{q\to 0} u'(q) = \infty$.[13] Similarly, $c(q)$ is $\mathcal{C}^2$ and satisfies $c'(q) > 0$, $c''(q) > 0$, $c(0) = 0$, $\lim_{q\to\infty} c'(q) = \infty$ and $\lim_{q\to 0} c'(q) = 0$. Therefore, it is socially optimal to

[11] q_t can also be interpreted as quality instead of quantity which might be more appropriate for some applications.

[12] We interpret y_t as net production of a borrower.

[13] The expectations operator captures the randomness implied by the matching process and the stochastic life expectancy.

produce q as $u(q) - c(q) > 0$ for some $q > 0$. In particular, let us define the *first best quantity* by q^* where $u'(q^*) = c'(q^*)$.[14]

When borrowers and lenders are matched during the BS, they have the opportunity to conduct the following trade: the lender produces q_t credit goods for the borrower on the spot and, as compensation, the borrower promises to produce b_t units of settlement goods for the lender in the subsequent RS. That is, borrowers finance their consumption of q_t by borrowing b_t from lenders, which is repayable by the end of the period. The terms of trade are determined by the borrower making a take-it-or-leave-it offer to the lender which the lender can either accept or reject. If the lender accepts, the credit good is produced. Otherwise the match is dissolved, no credit goods are produced and both parties proceed to the next RS.

We make the following assumptions on the information structure of the economy: all current and past actions are perfectly observable by anyone but there is limited knowledge about who committed those actions. To be precise, borrowers use an *account*[15], $a \in \mathcal{A}$ where $\mathcal{A}$ is any set with $|\mathcal{A}| = \infty$, when interacting with a lender and all actions undertaken by a given account are perfectly observable.[16] We denote any action x by x^a to denote that the action was committed with account a. Let us denote by $\xi_t^a = \{m_j^a, q_j^a, b_j^a, y_j^a\}_{j=0}^t$ the *history of past actions* which records all actions undertaken by account a up to period t where m_j^a is an indicator variable equal to one if the account was matched with a lender in period j and equal to zero if not. Importantly, however, the ownership of these accounts is private information and borrowers may create as many accounts as they wish. Borrowers can create new accounts during each RS at zero cost (we will modify this in section 5.4).

5.3 Costless accounts

We will now construct a particular equilibrium in recursive form. Let us first define *reputation* as a mapping from the account's history to a natural number. The reputation of account a at time t is denoted by n_t^a. We can then define reputation recursively: if an account has no history

[14] Because marginal cost and marginal utility are always equalised at every level of production for the settlement good, the social planner only cares about the production of the credit good.

[15] Equivalently, one could use the term *pseudonym*.

[16] We assume that accounts are unique. Hence, it is not possible to imitate another agent by taking on her history.

(i.e. $m_j^a = 0$ for all $j \leq t$), then $n_t^a = 0$ and

$$n_{t+1}^a = \begin{cases} \min\{n_t^a + 1, N\} & \text{if } m_t^a = 1 \text{ and } b_t^a \leq y_t^a \\ n_t^a & \text{if } m_t^a = 0, \\ -1 & \text{else.} \end{cases}$$

We say that an account is a *deviator* if $n_t^a = -1$ (more on that below). Intuitively, all accounts start with zero reputation and gain reputation by matching with lenders and repaying their debt. As the definition makes clear, we assume that there is some maximum level of reputation $N \in \mathbb{N}$ that can be achieved, so that $n_t^a \in [-1, 0, 1, \ldots, N]$. If accounts do not repay their debt, their account is marked as a deviator (i.e. $n_{t+1}^a = -1$).

Going forward, we want to construct an equilibrium with the following particular properties. First, agent's actions are conditioned only on reputation and not the entire history of actions of their accounts. Another way of saying this is that, in equilibrium, reputation is a sufficient statistic for the history of actions. We can therefore proceed without referring to the history of actions explicitly. Second, only pure-strategy equilibria are considered. Third, we assume that lenders do not lend to accounts which are marked as deviators.[17] This will serve two purposes: First it makes defaulting costly. Second it will introduce a punishment for "excessive borrowing" (i.e. borrowing more than the equilibrium amount). We will see later that this is crucial for having positive debt limits in equilibrium. Fourth, we restrict ourselves to equilibria where $b_{t,n} \geq b_{t,n-1}$ $\forall n > 0$. That is, in equilibrium a higher level of reputation will never decrease the amount that borrowers borrow. Moreover, we need to make the following assumption:

Assumption 1. *Borrowers maximally hold two accounts with* $n > 0$*.*

Let us briefly discuss the meaning, necessity and importance of this assumption. To be clear, we are not restricting how many accounts borrowers can use over their lifetime. Rather the restriction is on the simultaneous holding of more than two accounts with positive reputation. Therefore, we still allow borrowers to actively use multiple accounts. Using this assumption we are able to prove that in the equilibrium we study, borrowers have no incentive to use a second account even though they could. Naturally, it would be preferable to prove this more generally, but it also seems reasonable that if a borrower has no incentives to use a second account if she can use up to two, it would be very surprising to learn that this would change as soon as the

[17] Given that lenders believe that other lenders refuse to trade with deviators, it is in fact optimal for them to do so as well. The reason is straightforward: the borrower's incentive to honour their debt is to avoid being marked as a deviator. As a result, accounts already marked as deviators have "nothing to lose" and will therefore default on any promise. Lenders would anticipate this and therefore refuse to lend to a deviator.

borrower could use three or more accounts.

Finally, a couple of words on presentation. A borrower's individual state variables include the reputation of all her accounts. However, given the assumption that borrowers can hold only up to two accounts with $n > 0$ and the fact that borrowers can always replace deviator accounts with new accounts[18], we will omit the reference to any deviator accounts and simply write a borrower's state space consisting of the reputation of two accounts (all other accounts are either deviators or have zero reputation). Also, we will suppress time indexes unless unclear and denote any generic variable x by x_n to denote it being conditional on reputation n.

We proceed by describing the model in the following order: first, we start from the second subperiod, the RS, then, second, move on to describe the bargaining problem. Third, we discuss the BS-value function and finally, we define the equilibrium formally.

5.3.1 The repayment stage

Consider the problem of a borrower entering the RS with two accounts. Let us generally call these accounts 1 and 2, where each account has reputation (n^1, n^2) respectively. Let us adopt the convention that if we write the borrower's state variables as (n^1, n^2) we implicitly assume that the first entry corresponds to the account used in the last subperiod (so that we do not need to introduce an additional variable in the state space). Let us now denote the value function of a borrower who has matched with a lender in the previous subperiod by:

$$W_1^b(b, n^1, n^2) = \max_{\eta} -\eta b + \beta\Big[\eta V^b(\min\{N, n^1+1\}, n^2) + (1-\eta)V^b(0, n^2)\Big] \tag{5.1}$$

where $\eta \in \{0, 1\}$ is the optimal decision of a borrower with accounts (n^1, n^2) to repay its debt b and let us denote it's solution by $\eta(b, n^1, n^2)$. If they repay ($\eta = 1$), they incur linear costs to produce for lenders and their first account will gain reputation (unless reputation is already at maximum reputation). If they decide to default ($\eta = 0$), then the borrower does not suffer any disutility from producing the settlement good but the agent's first account will be marked as a deviator and, as explained in the previous section, the borrower then optimally replaces the account with a new one with no reputation such that $n^1 = 0$ in the next period. It is straightforward to see that $\eta(b, n^1, n^2) = 1$ if and only if the following *no-default* (ND) constraint is satisfied:

$$B(n^1, n^2) \equiv \beta\Big(V^b(\min\{N, n^1+1\}, n^2) - V^b(0, n^2)\Big) \geq b \tag{5.2}$$

[18]It is easy to see that since accounts are free to create, it is a (weakly) optimal strategy to create a new account as soon as an account has been marked as a deviator.

where $V^b(n^1, n^2)$ is the BS value function for a borrower with accounts (n^1, n^2). We define the *debt-limit* as the left-hand side of (5.2) which is the value above which a borrower with accounts (n^1, n^2) will default (i.e. $\eta = 0$). Hence, borrowers only repay their debt if the amount of debt is less or equal than the cost of losing the account's reputation.

Next, let us consider the case where borrowers did not match with any lenders in the previous subperiod. In that case, borrowers did not issue any debt and the value function can be written as:

$$W_0^b(n^1, n^2) = \beta V^b(n^1, n^2). \tag{5.3}$$

Similarly, we denote the RS-value function of a lender who enters the RS with claims on debts b issued by a borrower with an account with reputation n^1 by:

$$W^l(b, n^1) = b(1-p)\rho(b, n^1) + \beta V^l, \tag{5.4}$$

where $\rho(b, n^1)$ is the lender's belief that a borrower using an account with reputation n^1 decides to pay back its debt b. Note that because a borrower's reputation on its second account is unobservable to the lender, lenders cannot perfectly anticipate a borrowers default. They must therefore form expectations about the probability of default. In addition, lenders can perfectly anticipate that p borrowers will exit the economy and therefore not pay back their debt.

5.3.2 Terms of trade

The terms of trade are determined by the borrower making a take-it-or-leave-it offer to the lender. The offer takes the form (q, b), that is it specifies the production of credit goods by the lender, q, and the amount borrowed in terms of settlement goods by the borrower, b. The lender can either accept or reject. If the lender accepts, the credit good is produced and debt is recorded. If the lender rejects, the match is dissolved. A borrower with accounts (n_1, n_2) solves:

$$\begin{aligned}&\max_{q,b} u(q) + (1-p)\left[W_1^b(b, n^1, n^2) - W_1^b(0, n^1, n^2)\right]\\ &\text{s.t. } W^l(b, n^1) - W^l(0, n^1) = c(q),\end{aligned}$$

such that the borrower maximises her surplus taking into account the lender's participation constraint. Using the RS value functions (5.1) and (5.4), the problem can be rewritten as:

$$\begin{aligned} &\max_{q,b} u(q) - \eta(b, n^1, n^2)(1-p)b \\ &\text{s.t. } (1-p)\rho(q, n^1)b = c(q). \end{aligned} \tag{5.5}$$

Let us denote the resulting equilibrium values as q_n and b_n.

5.3.3 The borrowing stage

The value function for a borrower entering the BS with accounts (n^1, n^2) can be written as

$$\begin{aligned} V^b(n^1, n^2) =& \sigma \max\Big\{u(q_{n^1}) + (1-p)W_1^b(b, n^1, n^2), u(q_{n^2}) + (1-p)W_1^b(b, n^2, n^1)\Big\} \\ &+ (1-\sigma)(1-p)W_0^b(0, n^1, n^2). \end{aligned} \tag{5.6}$$

When agents enter the BS they either match with a lender or no match occurs. If they meet a lender they choose which account they use. As seen from the bargaining problem (5.5), the choice of account influences the terms of trade as different accounts may have different levels of reputation. Finally, observe that the borrower only proceeds to next period's RS if she does not exit the economy which occurs with probability $(1-p)$.

From equation (5.6) one can also infer that a borrower never uses a second account given the first account has positive reputation if and only if:

$$u(q_n) + (1-p)W_1^b(b_n, n, 0) > u(q_0) + (1-p)W_1^b(b_0, 0, n) \quad \forall n > 0. \tag{5.7}$$

In simple terms, if using an account with positive reputation yields more instantaneous utility from consuming and a higher continuation value compared to using an account with zero reputation, a second account is never used.[19]

The BS function for a lender can be written similarly:

$$V^l = \sigma \int_0^N \Big(-c(q_n) + W^l(b_n, n)\Big) dF(n) + (1-\sigma)W^l(0,0)$$

where $F(n) \in [0,1]$ is the *distribution of reputation*, i.e. $F(n)$ is the probability of matching with an account with reputation less or equal to n. Intuitively, if lenders match they encounter a random borrower with some given reputation n. Since borrowers search with accounts of

[19] At first, this may seem to be obvious since we assumed that more reputation allows to trade more. However, the issue is more subtle as borrowers could use the second account to eventually default on her promise while using the primary account as a "back-up". As we will see below however, in our equilibria this cannot occur.

varying reputation levels, the lender's counterparty possesses a stochastic reputation level with a known distribution function, $F(n)$. More generally, let us also define $G(n^1, n^2) \in [0,1]^2$ as the *distribution of accounts* where $G(n^1, n^2)$ is the probability that a random borrower has accounts with reputation equal or less to (n^1, n^2) respectively.[20] Let us also define their corresponding probability mass functions by $f(n)$ and $g(n^1, n^2)$ such that $F(n) = \sum_{\hat{n}=0}^{n} f(\hat{n})$ and $G(n^1, n^2) = \sum_{\hat{n}^1=0}^{n^1} \sum_{\hat{n}^2=0}^{n^2} g(\hat{n}^1, \hat{n}^2)$.

5.3.4 Equilibrium

Now we define the equilibrium. Specifically, we examine a *stationary*, *symmetric*, and *no-voluntary-default* equilibrium. Stationarity here means that $F(n)$ and $G(n^1, n^2)$ remain constant over time. Symmetry indicates that agents with identical state variables act in the same manner. No-voluntary default implies that borrowers always choose to repay when given the chance to do so on the equilibrium path (if borrowers randomly exit, they cannot repay).[21]

Definition 5 (Reputation equilibrium). *A stationary, symmetric and no-voluntary default equilibrium with a monotone reputation system and given maximum reputation N is given by a list of credit good consumption $\{q_n\}_{n=0}^{N}$, debts $\{b_n\}_{n=0}^{N}$ and debt limits $\{B_n\}_{n=0}^{N}$ such that*

1. *the RS-value functions are given by (5.1), and (5.3),*
2. *debt limits are determined by (5.2),*
3. *(q_n, b_n) solve the bargaining problem (5.5),*
4. *the borrower's BS-value function is given by (5.6),*
5. *borrowers always decide to repay, i.e. $\eta = 1$,*
6. *lender's believe that borrowers default on any off-equilibrium offers made, i.e. $\rho(b, n) = 1$ only if $b = b_n$.*

We proceed by characterising the equilibrium. First, given our assumption on the beliefs of lenders, we can rewrite the lender's participation constraint in the bargaining problem (5.5) to

$$b_n = \frac{c(q_n)}{(1-p)}. \tag{5.8}$$

[20] One can characterise $F(n)$ and $G(n^1, n^2)$ quite easily as they can be derived from an underlying law of motion based on the definition of reputation and the fact that agents stochastically exit the economy. However, as will soon become clear, knowing $F(n)$ is not necessary to characterise the equilibrium. Hence, we do not derive it explicitly.

[21] Our equilibrium definition is based on the concept of a perfect Bayesian equilibrium. That is, our equilibrium will be defined by strategies and beliefs such that strategies are sequentially rational and beliefs are derived, if possible, by Bayes rule. See Mas-Colell et al. (1995) for a formal definition.

That is, given lender's belief that borrowers will always repay if given the chance to do so, the lender's expected utility from receiving debt b is $b/(1-p)$ where $1/(1-p)$ is the default premium which arises from the fact that p borrowers exit exogenously. Therefore, the borrower's offer equalises the lender's cost with their expected benefit.

Second, since lenders belief that any-off equilibrium offer will not be repaid, the bargaining problem reduces to:

$$\max\Big\{u(q_n) - c(q_n), 0\Big\}.$$

Therefore, it must be that $q_n \leq \bar{q}$ for all n where $\bar{q}$ is given by $u(\bar{q}) = c(\bar{q})$. Since the surplus is increasing for any $q \leq q^*$, let us from now on assume that this is the case which implies that $q_n < \bar{q}$ for any n.

Third, from this we can use condition (5.2) and derive the following necessary condition for repayment to be always optimal:

$$\frac{c(q_n^1)}{1-p} \leq B(n^1, n^2) \tag{5.9}$$

for any (n_1, n_2) satisfying $g(n_1, n_2) > 0$. That is, for every distribution of accounts $g(n_1, n_2)$ that occurs in equilibrium borrowers need to be willing to repay their debt.

Observe that since lenders do not know the reputation of the borrower's second account, the following condition ensures repayment:

$$\frac{c(q_n^1)}{1-p} \leq \bar{B}(n^1) = \min_{n^2}\{B(n^1, n^2)\}. \tag{5.10}$$

This is a sufficient condition for (5.9) to hold. If condition (5.9) holds then any possible borrower with (n^1, n^2) will repay even if $g(n^1, n^2) = 0$ in equilibrium.

We can then show the following:

Lemma 7. *(5.7) is always satisfied if (5.10) holds.*

Proof. Given (5.10) borrowers never default. Given that we have assumed a monotone reputation system, we know that $b_0 \leq b_1, ..., \leq b_N$. Using (5.8) this implies $q_0 \leq q_1, ..., \leq q_N$. If $q_N \leq q^*$ then $S(n) = u(q_n) - c(q_n)$ is weakly increasing in n. The borrower's utility is then simply the discounted sum of $\sigma S(\hat{n})$ where $\hat{n}$ is the reputation of any account used. But then since $S(n)$ is weakly increasing in n and because the best strategy to increase n as fast as possible is to use only one account, it implies (5.7). □

That is to say, there is no incentive for borrowers to open a second account if debt limits are sufficiently tight so that default is not beneficial. If defaulting is never an option, then the borrower's best strategy is to use one account to accumulate reputation as fast as possible to increase the sum of discounted future surpluses $u(q) - c(q)$.

But now observe that if borrowers never hold a second account, i.e. $g(n^1, n^2) = 0$ for any $n^1 > 0$ and $n^2 > 0$, then 5.9 reduces to:

$$\frac{c(q_n)}{1-p} \leq B(n, 0). \tag{5.11}$$

Of course, so far we have only shown that (5.10) implies (5.7). Whether or not (5.11) also implies (5.7) is not a priori clear. However we can show that this is in fact the case.

Proposition 12. *If debt limits are given by (5.11) then borrowers only use one account, i.e. (5.11) implies (5.7).*

Proof. See appendix 5.6.1. □

The proposition implies that ensuring that borrowers with only one account are willing to repay, as specified in condition (5.11), also implies that borrowers never resort to use a second account. This may not be immediately apparent, as (5.11) does not imply that borrowers with multiple accounts are inclined to repay their debts. This raises the question of whether there exists a profitable deviation, where a borrower creates a second account and potentially defaults in the future. However, proposition 12 demonstrates that, although it might be advantageous for a borrower to default if they end up with multiple accounts, it is not beneficial to create a second account initially. The intuition behind this result is that regardless of whether a borrower intends to default or not, concentrating efforts on one account is the optimal strategy, leading to the accumulation of the most reputation and thus yielding the greatest benefits.[22]

Going forward, we can therefore simplify the notation by dropping any reference to the second account's reputation as, according to proposition 12, it is always zero (for example, $V^b(n^1, 0) = V^b(n)$).

Moreover, by combining (5.1), (5.3) and (5.6) and the fact that only one account is ever used, the BS-value function takes the following form:

$$V^b(n) = \sum_{j=0}^{\infty} \frac{(\beta(1-p)\sigma)^j}{(1-(1-\sigma)\beta(1-p))^{j+1}} \sigma S_{n+j}, \tag{5.12}$$

[22] The mechanism employed in Wang and Li (2023) to prevent multiple accounts being used in equilibrium is different. They assume that accounts are punished if the account is observed to be inactive for one period. In their environment borrowers and lenders always match and therefore observing that an account was not used implies that another account was used. Such a mechanism would not work here as borrowers and lenders do not always match. Therefore, observing that an account did not trade could either be because a borrower did not match with a lender or because a different account was used.

where $S_n \equiv u(q_n) - c(q_n)$ for all $n \leq N$ and $S_n = S_N$ for all $n > N$.

Lemma 8. *Any sequence $\{q_n\}_{n=0}^N$ which satisfies (5.11) and (5.12) is a reputation equilibrium.*

Proof. Proposition 12 indicates that (5.11) implies (5.7). Furthermore, (5.7) implies $g(n^1, n^2) = 0$ for any $n^1 > 0$ and $n^2 > 0$. Given this, we immediately see that (5.9) implies (5.11). Finally, the value function (5.1), (5.3) and (5.6) can be collapsed into equation (5.12). □

Proposition 13. *In a reputation equilibrium $q^* = q_n$ for all n cannot be an equilibrium.*

Proof. Suppose $q^* = q_n$ for all n is an equilibrium. Then by (5.12) it follows that $V^b(n) = V^b$ for all n. But then according to (5.11) $c(q^*) = 0$ which is a contradiction. □

This is not so surprising as trading the first best quantity q^* at every level of reputation implies that reputation conveys no benefits and therefore the loss of reputation, the punishment for defaulting, has no effect. As a result, the equilibrium cannot be sustained.

Going forward, it is useful to differentiate between two types of equilibria.

Definition 6 (**Not-too-tight and too-tight equilibria**). *A reputation equilibrium is called not-too-tight if (5.11) holds with equality for all n. Otherwise, we call the equilibrium too-tight.*

The "not-too-tight" terminology originates from Alvarez and Jermann (2000) and implies that $b_n = B_n$ for all n. That is, if an equilibrium is not-too-tight then the borrower is indifferent between repaying and defaulting for each level of reputation. Therefore, in each match, lenders lend the maximum amount of debt such that the borrower is willing to repay. This is the most natural equilibrium to study as it maximises the gains from trade between a borrower and a lender while ensuring that borrowers always repay.

5.3.4.1 Not-too-tight equilibria

Let us for now assume that $b_n = B_n$ for all n. We can show the following:

Proposition 14. *In a not-too-tight equilibrium it must be that $q_n = 0$ for all n.*

For illustrative purposes we assume $\sigma = 1$ and $N = 1$.

Proof. Consider the no-default constraints (5.2) for $n = 0$ and $n = 1$

$$B_0 = \beta[V^b(1) - V^b(0)] \geq \frac{c(q_0)}{(1-p)}, \tag{5.13}$$

$$B_1 = \beta[V^b(1) - V^b(0)] \geq \frac{c(q_1)}{(1-p)}, \tag{5.14}$$

which imply that $B_1 = B_0$. Given we study a not-too-tight equilibrium, this implies $b_0 = b_1$ and from (5.8) it follows that $q_0 = q_1$. But then $V^b(1) = V^b(0)$ and $B_1 = B_0 = 0$ which implies $q_0 = q_1 = 0$. □

A general proof can be found in the appendix 5.6.2. The intuition for this result can be understood in the following way: in the simplified case where $N = 1$ borrowers either have reputation ($n = 1$) or they have none ($n = 0$). In this case, a borrower's debt limit depends on the value of reputation, $V^b(1) - V^b(0)$, which is independent of the borrower's current level of reputation (see equations (5.13) and (5.14)). Therefore, a borrower with and a borrower without reputation has the same debt limit. Since in a not-too-tight equilibrium borrowers borrow up to the maximum level of debt such that they are willing to repay, it then implies that borrowers borrow the same amount irrespective whether they have reputation or not. But this is problematic because this makes reputation not valuable in the first place since the amount of trade for a borrower with $n = 0$ is the same as for a borrower with $n = 1$. If reputation has no value, there is no incentive to repay and lenders are thus not willing to accept any amount of debt.

How should we interpret this result? A valid interpretation of the not-too-tight equilibrium is that it is the only surviving equilibrium if we apply the intuitive criterion by Cho and Kreps (1987). That is to say, as long as an offer satisfies (5.11), the lender should not believe that the borrower will default. However, since $q < q^*$, the borrower would optimally borrow as much as possible, and the lender is willing to accept, as the condition (5.11) is satisfied, ensuring that the borrower will always repay. Therefore, the results suggest that for the equilibrium to exist, lenders need a reason to reject off-equilibrium offers even if those offers satisfy (5.11).

In our approach going forward, we will be agnostic about the exact reason why lenders would do so and stick to the definition 5 (and hence the concept of a Perfect-Bayesian equilibrium). However, it is not hard to see that a slightly modified equilibrium definition would provide such a reason. For example, if we mark borrowers as deviators not only for defaulting but also for making off-equilibrium offers, lenders would reject such offers even given the intuitive criterion. This is because borrowers would default on any off-equilibrium offer since they will be marked as deviators anyway.

5.3.4.2 Too-tight equilibria

From now on, we will focus on too-tight equilibria. Thus, it must be that (5.11) is not holding with equality for at least one n. We proceed by first defining an optimal reputation equilibrium, then derive sufficient conditions for such an optimal reputation equilibrium to exist and then finally show that this optimal reputation equilibrium always exists (naturally, this also shows that reputation equilibria always exist).

We start by defining an optimal reputation equilibrium:

Definition 7 (**Optimal reputation equilibrium**). *An optimal reputation equilibrium solves*

$$\max_{\{q_n\}_{n=0}^N} V^b(0) \ \textit{s.t.} \ V^b(\min\{N, n+1\}) - V^b(0) \geq \frac{c(q_n)}{\beta(1-p)} \quad \forall n \textit{ where } V^b(n) \textit{ is given by } (5.12). \tag{5.15}$$

Observe that an optimal reputation equilibrium is a reputation equilibrium which maximises a borrower's lifetime utility so that the borrower repays.[23] We can then derive the following proposition:

Proposition 15. *A sufficient condition for an optimal reputation equilibrium is* $V^b(0) = \bar{V}(\hat{q})$ *where* $\hat{q}$ *and* $\bar{V}(\hat{q})$ *are determined by*

$$\frac{u'(\hat{q})}{c'(\hat{q})} = 1 + \frac{1-\beta(1-p)}{\sigma\beta(1-p)}, \tag{5.16}$$

$$\bar{V}(\hat{q}) = \frac{\sigma[u(\hat{q}) - c(\hat{q})]}{1-\beta(1-p)} - \frac{c(\hat{q})}{\beta(1-p)}. \tag{5.17}$$

Proof. Using the ND-constraint for $n = N$, inserting (5.12) for $n = N$ and solving for $V^b(0)$ yields $V^b(0) \leq \bar{V}(q_N)$ where $\bar{V}(q_N)$ is given by (5.17). The upper bound, $\bar{V}(q_N)$, depends solely on q_N. The first order condition of (5.17) with respect to q_N yields (5.16). If $V^b(0) = \bar{V}(\hat{q})$ then the equilibrium must solve (5.15) because any $V^b(0) > \bar{V}(\hat{q})$ can never satisfy (5.12) for $n = N$ given that $\bar{V}(\hat{q})$ is maximal at $q_N = \hat{q}$. □

Hence, $\bar{V}(\hat{q})$ is the highest value of $V^b(0)$ that can be achieved and if we find a sequence $\{q_n\}_{n=0}^N$ that achieves $\bar{V}(\hat{q})$ we have found an optimal reputation equilibrium. Of course, it is not a priori clear whether the upper bound $\bar{V}(\hat{q})$ can be attained throughout the entire parameter space. Let us now consider the kind of sequences $\{q_n\}_{n=0}^N$ that achieve $\bar{V}(\hat{q})$. To gain some intuition, it is useful to consider the simplified case where $N = 1$ and $\sigma = 1$ (again, accounts either have reputation or they have none). The ND-constraints for $n = 1$ simplifies to:

$$V_1^b - V_0^b = S_1 - S_0 = \frac{c(q_1)}{\beta(1-p)}, \tag{5.18}$$

while the ND-constraint for $n = 0$ is satisfied if $q_0 \leq q_1$. The sufficient condition derived in

[23]The lender's expected lifetime value is always zero because the borrower, by making a take-it-or-leave-it offer, appropriates all the surplus.

proposition 15 implies that optimally $q_1 = \hat{q}$ and q_0 is set as such that (5.18) holds. If the implied q_0 is between zero and q_1 the equilibrium exists. From (5.18) we can easily see that $q_0 \leq q_1$ is always satisfied as $S_1 > S_0$ requires $q_1 > q_0$. On the other hand, $q_0 \geq 0$ holds if and only if $S_0 \geq 0$ which, according to (5.18), is the case if:

$$u(\hat{q}) \geq c(\hat{q})\frac{1+\beta(1-p)}{\beta(1-p)}. \tag{5.19}$$

These equations have a clear interpretation: equation (5.18) tells us that the value of having reputation, $V_1^b - V_0^b$, is given by higher surpluses which can be obtained with reputation, $S_1 - S_0 > 0$, and this value has to optimally equal the gain of defaulting, $\frac{c(q_1)}{\beta(1-p)}$. Moreover, even though $q_1 = q^*$ might be feasible, it is in general not optimal. The reason is that trading q^* would require a high level of reputation to incentivise borrowers to not default. This in turn is only achievable if q_0 is sufficiently small. However, due to the concavity of $u(q)$ it is optimal to set $q_1 < q^*$ in order to increase q_0. Finally, the upper bound can only be achieved if (5.19) is satisfied. Intuitively, the highest value of reputation is attained if $q_0 = 0$. For some parameter values however, even this value is not sufficient to incentivise agents to repay $\hat{q}$. However, as we show next, one can increase N to ensure that the upper bound can be achieved.

In the more general case where $N > 1$ we find that an optimal reputation equilibrium can be implemented with infinitely many sequences. This indeterminacy follows from the fact that in order to achieve the upper limit $\bar{V}(\hat{q})$ only two equations need to hold with equality: $q_N = \hat{q}$ and the ND-constraint for $n = N$. The other ND-constraints do not need to hold with equality. Thus, there are $N+1$ variables to determine and two equations to pin them down. One way to find possible solutions is to compute them numerically. Figure 5.1 plots several sequences with different N which all achieve $\bar{V}(\hat{q})$.[24]

There are however also closed form solutions. Consider the following: set $q_N = \hat{q}$ and $q_n = 0$ for all $n < N-1$. Notice, that the ND-constraint for all $n < N-1$ are in that case automatically satisfied. The implied solution for q_{N-1} must however satisfy $0 \leq S_{N-1}$ (no negative surplus possible) and $S_{N-1} < S_N$ (otherwise the ND-constraint for $n = N-1$ would be violated). We can show the following:

Proposition 16. *There always exists a reputation equilibrium such that $V^b(0) = \bar{V}(\hat{q})$ for a particular $N = \hat{N}$.*

Proof. See appendix 5.6.3. □

[24]We assumed the following functional forms: $u(q) = \frac{q^{1-\eta}}{1-\eta}$ and $c(q) = q$. The parameters that were used are: $\beta = 0.96$, $\eta = 0.5$, $\sigma = 0.5$ and $p = 0.1$.

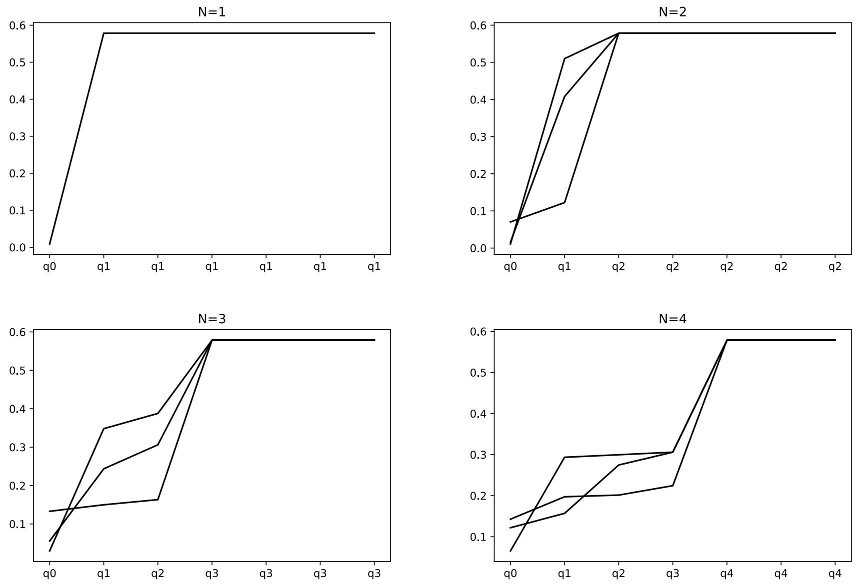


Figure 5.1: Different equilibria that achieve $\bar{V}(\hat{q})$

In the proof we derive $\hat{N}$ analytically. As we have seen in the special case where $N = 1$ and $\sigma = 1$, there was the possibility that the harshest possible punishment (setting $q_0 = 0$) was not sufficient to guarantee that agents repay their debt associated with trading $\hat{q}$. The proposition indicates that we can always increase N and thereby increase the scope for punishment in order to guarantee that an equilibrium always exists.

In conclusion, we have shown that credit can, in principle, be anonymous and not be reliant on any form of collateral. According to proposition 16 such credit equilibria do always exist provided that N is chosen appropriately. However, while such credit equilibria always exist proposition 15 indicates that they are costly in the sense that the punishment for default implicitly is imposed on young agents too. There is therefore a trade-off between making defaulting costly and letting young agents consume.

5.4 Costly accounts

So far we have assumed that borrowers can create accounts for free. We now modify our original model by making accounts costly to create. We believe this modification to be highly practically relevant as its implementation is simple. In fact, there are already many providers of blockchain wallets which ask for a fee in order to set up a new wallet.[25]

[25]For example to open up the wallets Trezor Model T and Ledger Nano X a user pays \$100-\$300 (source: CNET Money).

5.4.1 Modification to the baseline model

We make three modifications to the environment. First, each time an account is opened up in the RS, the borrower is required to pay a *fee* $\kappa \in \mathbb{R}^+$ in terms of settlement goods. We consider two cases: In the first case the fee is considered a real cost and therefore a dead-weight loss. In the second case, the fee is viewed to be a transfer to some benevolent authority managing the accounts and which redistributes those fees (for example, this could be the government or a private platform). Let us call the redistributed fees *transfers* and denote the per-capita transfer by τ. In order to nest both of these cases into one, let us assume that only a share $\epsilon \in [0, 1]$ of these costs can be redistributed as transfers and we will study the two polar cases.

Second, we modify the recording technology. To be specific, let us redefine the history of actions to be $\xi_t^a = \{m_j^a, q_j^a, b_j^a, y_j^a, k_j^a\}_{j=0}^t$ where k_j^a is an indicator variable which equals one if the fee was paid by account a in period j.

Third, we assume that accounts are deleted if the owner exits the economy. One could easily modify the model to make this an outcome. For example, assume that accounts need to pay a tiny fee each period in order to keep the account active. However, to keep notation to a minimum we introduce this as a restriction in the environment.

In addition to these changes in the environment, we will construct a different equilibrium compared to the case without fees. Let us therefore re-define the reputation n_t^a to be:

$$n_t^a = \begin{cases} 1 & \text{if } k_j^a = 1 \text{ for some } j < t, \text{ and } b_t^a \leq y_t^a, \\ -1 & \text{else.} \end{cases}$$

Let us call accounts with $n = 1$ to be *active accounts* and, as before, accounts with $n = -1$ to be deviator accounts. That is, an account is active if the account has paid the fee sometime in the past and the account has always repaid it's debts. Otherwise, an account is considered a deviator. We maintain the same assumption concerning the lender's beliefs: lenders will never trade with deviator accounts (i.e $n = -1$) and reject any off-equilibrium offers, i.e. $\rho(b, n) = 1$ only if $n = 1$ and $b = b_1$. But different to before, there is no longer any notion of increasing one's reputation over time. Moreover, let us assume that all active accounts receive transfers τ.

We will study an equilibrium where borrowers will only ever use one account and always repay endogenously. While we do not prove this formally, it should be clear that the reasoning behind proposition 12 should apply to this case too. Therefore, we do not need to keep track of multiple accounts as we did in the previous section.

5.4.2 Equilibrium

We first look at the RS-value function of a borrower who owns an active account ($n = 1$) who has issued b debt:

$$W_1^b(b) = \max_{\eta_b, \eta_k} \ \eta_b(\tau - b) - \eta_k \kappa + \beta\Big(\mathbb{I} V_1^b + (1 - \mathbb{I}) V_{-1}^b\Big) \tag{5.20}$$

where $\eta_b \in \{0, 1\}$ and $\eta_k \in \{0, 1\}$ are both discrete choices indicating whether the borrower repays it's debt or creates a new account respectively. Let us denote the solutions by $\eta(b)$ and $\eta(k)$ respectively. Furthermore, let us denote $\mathbb{I} = \mathbb{1}(\eta_b = 1 \vee \eta_k = 1)$ which indicates whether the borrower will enter the next period with an active account. Similarly, for a borrower with an inactive account:

$$W_{-1}^b(0) = \max_{\eta_k} -\eta_k \kappa + \beta\Big(\mathbb{I} V_1^b + (1 - \mathbb{I}) V_{-1}^b\Big). \tag{5.21}$$

From (5.21) one can easily see that $\eta_k = 1$ if and only if:

$$\kappa \leq \beta(V_1^b - V_{-1}^b). \tag{5.22}$$

Clearly, this is a necessary condition for an equilibrium to exist because if this condition is not satisfied borrowers would never hold an active account. Furthermore, from equation (5.20), we can conclude that borrowers will repay their debt ($\eta(b) = 1$) if and only if the following no-default constraint is satisfied (conditional on 5.22 holding):

$$B(\kappa, \tau) \equiv \kappa + \tau \geq b \tag{5.23}$$

where, similar to the previous section, $B(\kappa, \tau)$ is the debt limit which corresponds to the maximum amount of debt which the borrower is willing to repay. We observe from the no-default constraint (5.23) that borrowers only repay their debt if the amount of debt minus transfers is less or equal than the cost of setting up a new account. Notice, that both a higher fee κ and a higher transfer τ increases the debt limit (the latter because transfer are only received by active accounts). Naturally, in equilibrium, the no-default constraint has to be satisfied in order that borrowers always repay their debt, i.e. $\eta(b) = 1$.

Next, we write the bargaining problem between a lender and a borrower. The same way as before, the borrower makes a take-it-or-leave-it offer:

$$\max_{q,b} u(q) - (1 - p)b$$
$$\text{s.t. } b(1 - p)\rho(b, n) = c(q).$$

Given our assumption on lender's belief we again find that:

$$b(1-p) = c(q) \tag{5.24}$$

for any equilibrium offer $b = b_1$.

Furthermore, the BS-value function of the borrowers with an active and deviator account are, respectively, given by:

$$V_1^b = \sigma[u(q_1) + (1-p)W_1^b(b_1)] + (1-\sigma)(1-p)W_1^b(0), \tag{5.25}$$

$$V_{-1}^b = (1-p)W_{-1}^b(0). \tag{5.26}$$

The same holds for the RS- and BS-value functions of the lenders which are given by $W^l(b_1) = (1-p)b_1 + \beta V^l$ and $V^l = \sigma[-c(q_1) + W^l(b_1)] + (1-\sigma)W^l(0)$.

In equilibrium the amount of paid fees must equal the amount of redistributed fees in each period (as settlement goods cannot be stored). As the mass of agents paying the fee is p (new entrants) and the one receiving it is $(1-p)$ (all agents without new entrants) it follows that

$$\epsilon \cdot p \cdot \kappa = (1-p)\tau. \tag{5.27}$$

We can now define the equilibrium formally.

Definition 8 (**Costly accounts equilibrium**). *A stationary, symmetric and no-voluntary-default equilibrium with costly accounts is given by credit good consumption q_1, debts b_1, debt limits B_1 and transfers τ such that*

1. *the value functions are given by (5.20), (5.21), (5.25) and (5.26),*
2. *borrowers are willing to open accounts, (5.22),*
3. *borrowers always decide to repay, (5.23),*
4. *q_1 and b_1 satisfy (5.24),*
5. *the size of the transfer is determined by (5.27)*

for any κ.

Next, we can insert (5.24) and (5.21) into (5.25):

$$V_1^b = \frac{\sigma S_1 + (1-p)\tau}{1-\beta(1-p)}$$

where $S_1 \equiv u(q_1) - c(q_1)$. Similarly, we insert (5.24) and (5.20) into (5.26):

$$V_{-1}^b = -\kappa(1-p) + \beta(1-p)\frac{\sigma S_1 + (1-p)\tau}{1-\beta(1-p)}.$$

Using these two equations we find that:

$$V_1^b - V_{-1}^b = \sigma S_1 + (1-p)(\kappa + \tau) \tag{5.28}$$

which tells us that the value of having an active account is the ability to borrow with some probability, σS_1, the value of transfers received with probability $(1-p)$ in the next superiod, $(1-p)\tau$, and since fees only have to be paid once, the value of the fee paid with probability $(1-p)$ in the next subperiod, $(1-p)\kappa$. In a next step we prove existence of the costly accounts equilibrium.

Proposition 17. *A sufficient condition for a costly account to exist is:*

$$\bar{\kappa}(q_1) \equiv \frac{\beta\sigma(u(q_1) - c(q_1))}{1-\beta(1-p(1-\epsilon))} \geq \kappa \geq \frac{c(q_1)}{1-p(1-\epsilon)} \equiv \underline{\kappa}(q_1) \tag{5.29}$$

for any κ and q_1. Furthermore, there always exists some $q_1 > 0$ such that $\bar{\kappa}(q_1) > \underline{\kappa}(q_1)$.

Proof. We can derive the lower bound by combining (5.23) with (5.24) and (5.27):

$$\kappa \geq \frac{c(q_1)}{1-p(1-\epsilon)} \equiv \underline{\kappa}(q_1).$$

Furthermore, the upper bound is found by inserting (5.27) and (5.28) into (5.22):

$$\bar{\kappa}(q_1) \equiv \frac{\beta\sigma(u(q_1) - c(q_1))}{1-\beta(1-p(1-\epsilon))} \geq \kappa.$$

Next, we want to show that there always exists some q_1 for which $\bar{\kappa}(q_1) > \underline{\kappa}(q_1)$. Let us first define $\Omega(q_1) \equiv \bar{\kappa}(q_1) - \underline{\kappa}(q_1)$. Then observe that $\Omega(0) = 0$ and

$$\Omega'(q_1) = \frac{\beta\sigma(u'(q_1) - c'(q_1))}{1-\beta(1-p(1-\epsilon))} - \frac{c'(q_1)}{1-p(1-\epsilon)}.$$

But then observe that $\lim_{q_1 \to 0} \Omega'(q_1) = \infty$ because $\lim_{q \to 0} u'(q) = \infty$ and $\lim_{q \to 0} c'(q) = 0$. This implies that there exists some $q_1 > 0$ such that $\bar{\kappa}(q_1) > \underline{\kappa}(q_1)$. □

The proposition tells us that the fee κ can neither be too large or too small. If the fee is too low then borrowers have an incentive to default on their debt. However, if the fee is too large then borrowers never create an account in the first place. Of course, the feasible range for κ

depends on q_1 and, as we show, there always exists some q_1 so that there exists a feasible fee to incentivise borrowers to both create an account and repay their debts.

It is also easy to see that the equilibrium set is larger in the case where we interpret the fee not as a real cost, $\epsilon > 0$. In particular, going from $\epsilon = 0$ to $\epsilon = 1$ increases the upper bound and decreases the lower bound. Intuitively, being able to redistribute the fees increases the lifetime value of having an account (upper bound) and decreases the incentive to default (lower bound).

Finally, and again assuming that the fee can be redistributed, the result implies that a benevolent authority can always find some q_1 and associated κ that incentivises borrowers to participate and repay their debts. In the next section, we will ask: what is the optimal fee that a benevolent authority should charge?

5.4.3 Optimal costly accounts equilibrium

So far, we have shown that there are many combinations of q_1 and κ that constitute an equilibrium. We now want to show that there exists a uniquely optimal combination of q_1 and κ. As a result, from now on we will stick with the interpretation that the fee is not a real cost (i.e. $\epsilon > 0$) and can therefore be set by the benevolent authority. Let us then define the following:

Definition 9 (**Optimal costly accounts equilibrium**). *An optimal costly accounts equilibrium is given by q_1 and κ that maximises V_1^b subject to (5.27) and (5.29).*

That is, an optimal costly accounts equilibrium maximises the lifetime value of a borrower that newly enters the economy. The associated program is given by:

$$\max_{\kappa, q_1} -\kappa(1-p) + \beta(1-p)\frac{\sigma(u(q_1) - c(q_1)) + (1-p)\tau}{1-\beta(1-p)},$$

$$\text{s.t.}$$

$$\tau = \frac{p}{1-p}\epsilon\kappa,$$

$$\frac{\beta\sigma(u(q_1) - c(q_1))}{1-\beta(1-p(1-\epsilon))} \geq \kappa \geq \frac{c(q_1)}{1-p(1-\epsilon)}.$$

Let us denote the optimal quantity of credit good by $\tilde{q}$ and the optimal fee by $\tilde{\kappa}$. We assume that the benevolent authority redistributes all fees, i.e. $\epsilon = 1$.

Proposition 18. *The optimal costly accounts equilibrium is given by:*

$$\tilde{\kappa} = \frac{c(\tilde{q})}{1-p(1-\epsilon)},$$

$$\frac{u'(\tilde{q})}{c'(\tilde{q})} = 1 + \frac{1-\beta}{\sigma\beta}. \tag{5.30}$$

Proof. See appendix 5.6.4. □

One immediately observes that the optimal costly accounts equilibrium implies a credit good quantity which is below the first-best quantity, i.e. $\tilde{q} < q^*$. This is true even though q^* might be attainable. The intuition is similar to the reputation equilibrium: because the fee has to be financed in advance to borrowing and since agents are impatient, it is optimal to reduce the amount of borrowing by agents which allows account fees to be lowered. Another way of putting it is that at $q = q^*$ reducing q has a first order effect on κ but only a second order effect on $u(q) - c(q)$.

Furthermore, by comparing the optimal quantity of the costly accounts equilibrium (5.30) with the quantity consumed by a full reputation borrower ($n = N$) from the previous section, (5.16), one can immediately see that the amount consumed is strictly higher in the optimal costly accounts equilibrium.

5.5 Closing remarks

In this paper, we present a novel perspective on credit and anonymity, departing from the existing body of literature which either neglects agents' anonymity concerns entirely or imposes strict anonymity prerequisites that effectively rule out the existence of credit systems. Instead, our focus lies on pseudonymity, a particularly prevalent form of anonymity currently observed in many areas of the internet and, in particular, blockchains.

What can we learn from this exercise? First and foremost, there is often an assumption that anonymity and credit cannot coexist. Our analysis, from an economic standpoint, challenges this notion. We believe our rationale extends to credit in a broader context. This holds particular significance for blockchain-related endeavours aiming to integrate credit within blockchain networks. Secondly, our analysis shows that there are also costs in maintaining anonymous credit. The root of the costs is the impossibility to distinguish between first time entrants and entrants due to former default. This implies that if credit is at entrance kept low to punish former defaulters this also applies to first time entrants.

Lastly, we wish to address some potential concerns arising from our framework. Firstly, as our study explores merely a subset of all possible credit equilibria, we refrain from making definitive assertions about the optimality of credit systems. However, we conjecture that our analysis did indeed capture the most efficient credit equilibria. Secondly, a question arises regarding the significance of the assumption that borrowers only engage with one lender per period. It is evident that if borrowers were free to interact with an unlimited number of lenders, a dominant strategy would involve creating an infinite number of accounts, thereby reducing the debt limit to zero. While our current framework does not explicitly demonstrate this, we hypothesise that any cost of contacting additional lenders would suffice to prevent such a scenario. Thirdly, it is worth noting that our model simplifies many real-world complexities associated with credit. For example, we ignore potential heterogeneity in agents' ability to repay their debt. We leave all these considerations for future research.

5.6 Appendix

5.6.1 Proof of proposition 12

Proof. We show that (5.7) holds. In other words, we want to prove that it is never advantageous for a borrower to open up and gain reputation on a second account. To establish a contradiction, assume that the borrower opens a second account at time t_0, while already possessing a first account with reputation n^1. For brevity, we denote $\tilde{\beta} \equiv \beta(1-p)$. There are two primary scenarios concerning the reputation on the first account when the borrower opens the second account:

1. There exists an $n < n^1$ such that $q_{n^1} > q_n$. There are four feasible strategies which cover all possibilities:

   (a) Consider the case where the borrower opens a second account and accumulates reputation until it reaches $n^2 = n^1 + 1$. At this point, the borrower's state variables can be expressed as $(n^1, n^1 + 1)$, assuming that this occurs in some period t_1. The *deviation value* can be formulated as follows:

   $$\tilde{V} = \mathcal{S} + \tilde{\beta}^{t_1 - t_0 - 1}[u(q_{n^2-1}) - c(q_{n^2-1})] + \tilde{\beta}^{t_1 - t_0} V(n^1, n^1 + 1)$$

   where $\mathcal{S}$ is the discounted payoff of using the second account between t_0 and $t_1 - 1$.[26] In t_0 there is another feasible strategy: alternatively, the agent could use the first account one more time in t_0 and then start using the second account so that in t_1 the agent's state variables are $(n^1 + 1, n^1)$. This *alternative value* is then given by:

   $$V^* = [u(q_{n^1}) - c(q_{n^1})] + \tilde{\beta}\mathcal{S} + \tilde{\beta}^{t_1 - t_0} V(n^1 + 1, n^1).$$

   The deviation value can only be optimal if it weakly exceeds the alternative value: $V^* \leq \tilde{V}$. Inserting the values and recognizing that $q_{n^1} = q_{n^2-1}$ and $V(n^1 + 1, n^1) = V(n^1, n^1 + 1)$ implies

   $$[u(q_{n^1}) - c(q_{n^1})](1 - \tilde{\beta}^{t_1 - t_0 - 1}) \leq \mathcal{S}(1 - \tilde{\beta}).$$

   Since there exists an $n^2 \leq n^1$ such that $q_{n^1} > q_{n^2}$ it must be that $\mathcal{S} < [u(q_{n^1}) - c(q_{n^1})] \sum_{j=0}^{t_1 - t_0 - 1} \tilde{\beta}^j = [u(q_{n^1}) - c(q_{n^1})] \frac{1 - \tilde{\beta}^{t_1 - t_0}}{1 - \tilde{\beta}}$ and therefore the previous inequality

[26] The explicit expression is given by $\mathcal{S} = \sum_{n=0}^{n^2-2} \tilde{\beta}^{\hat{t}_n - t_0}[u(q_n) - c(q_n)]$ where $\hat{t}_n$ is the time where the borrower trades with the second account with reputation n.

implies: $u(q_{n^1}) - c(q_{n^1}) < u(q_{n^1}) - c(q_{n^1})$. This is a contradiction.

(b) Suppose the borrower opens a second account and builds up reputation until $n^2 \leq n^1$. He then switches accounts, borrows and repays with the first account so that the agent's state variables are (n^1+1, n^2) at this point. Without loss of generality, let us suppose the switch occurs in period t_1. The deviation value can be written as

$$\tilde{V} = \mathcal{S} + \tilde{\beta}^{t_1-t_0}[u(q_{n^1}) - c(q_{n^1})] + \tilde{\beta}^{t_1-t_0+1}V(n^1+1, n^2) \tag{5.31}$$

where $\mathcal{S}$ is the discounted payoff of using the second account between t_0 and $t_1 - 1$.[27] In t_0 there is another feasible strategy: alternatively, the agent could use the first account one more time in t_0 and then start using the second account so that the borrower's state variables are the same in t_1+1. This alternative value is then given by:

$$V^* = [u(q_{n^1}) - c(q_{n^1})] + \tilde{\beta}\mathcal{S} + \tilde{\beta}^{t_1-t_0+1}V(n^1+1, n^2).$$

The deviation value can only be strictly beneficial if it exceeds the alternative value: $V^* \leq \tilde{V}$. Inserting the values yields:

$$[u(q_{n^1}) - c(q_{n^1})](1 - \tilde{\beta}^{t_1-t_0}) < \mathcal{S}(1 - \tilde{\beta}).$$

Since there exists an $n^2 \leq n^1$ such that $q_{n^1} > q_{n^2}$ it must be that $\mathcal{S} < [u(q_{n^1}) - c(q_{n^1})]\frac{1-\tilde{\beta}^{t_1-t_0-1}}{1-\tilde{\beta}}$ and therefore the previous inequality implies: $u(q_{n^1}) - c(q_{n^1}) < u(q_{n^1}) - c(q_{n^1})$. This is a contradiction.

(c) Suppose the borrower opens a second account and builds up reputation until $0 \leq n^2 \leq n^1$. At some point, the agent defaults on the second account, and without loss of generality, let us assume this default occurs in period t_1:

$$\tilde{V} = \mathcal{S} + \tilde{\beta}^{t_1-t_0}u(q_{n^2}) + \tilde{\beta}^{t_1-t_0+1}\tilde{V}.$$

Importantly, the borrower's state variables and therefore the continuation value are the same before using the second account and defaulting on the second account. However, this would imply that the first account would never be used again and $\tilde{V} = V(n^1, 0)$. But since $\tilde{V} = V(n^1, 0)$ is a possible strategy with state variables $(0,0)$, it must be that $V(0,0) \geq V(n^1, 0) = \tilde{V}$. But according to the ND-constraint

[27] The explicit expression is given by $\mathcal{S} = \sum_{n=0}^{n^2-1} \tilde{\beta}^{\hat{t}_n - t_0}[u(q_n) - c(q_n)]$ where $\hat{t}_n$ is the time where the borrower trades with the second account with reputation n.

(5.11) this implies $q_n = 0$ for all $n < n^1$ and therefore $V(0,0) < V(n^1, 0) = \tilde{V}$. This is a contradiction.

(d) Suppose the borrower opens a second account and builds up reputation until $n^2 \leq n^1$. Then the agent switches back to the first account, borrows from it, and immediately defaults on that account. Without loss of generality, let us suppose the switch and default occur in period t_1. The deviation value can then be written as:

$$\tilde{V} = \mathcal{S} + \tilde{\beta}^{t_1 - t_0} u(q_{n^1}) + \tilde{\beta}^{t_1 - t_0 + 1} V(0, n^2).$$

Observe, that in t_0 there is another feasible strategy: the agent defaults in t_0 on the first account and then starts using the second account so that in t_1 the agent's state variables are the same as in $t_1 + 1$. The alternative value is then given by:

$$V^* = u(q_{n^1}) + \tilde{\beta}\mathcal{S} + \tilde{\beta}^{t_1 - t_0 + 1} V(0, n^2).$$

The deviation value can only be strictly beneficial if it weakly exceeds the alternative value: $V^* \leq \tilde{V}$. Inserting the values yields:

$$u(q_{n^1})(1 - \tilde{\beta}^{t_1 - t_0 - 1}) < \mathcal{S}(1 - \tilde{\beta}).$$

Since there exists an $n^2 \leq n^1$ such that $q_{n^1} > q_{n^2}$ it must be that $\mathcal{S} \leq [u(q_{n^1}) - c(q_{n^1})]\frac{1 - \tilde{\beta}^{t_1 - t_0 - 1}}{1 - \tilde{\beta}}$ and therefore the previous inequality implies: $0 < -c(q_{n^1})$. This is a contradiction.

2. There exists no $n < n^1$ such that $q_{n^1} > q_n$. Let us therefore denote $\hat{n}$ the lowest level of reputation such that $q_{\hat{n}} > q_n$ for all $n < \hat{n}$.

(a) The borrower trades with both accounts until the reputation of both accounts is given by $(\hat{n}_1, \hat{n}_2)$ where $\hat{n}_1 < \hat{n}$ and $\hat{n}_2 < \hat{n}$. The borrower then defaults on either of these accounts. Without loss of generality, let us suppose the default occurs on the second account and the switch and default occur in period t_1. The deviation payoff can be written as

$$\tilde{V} = \mathcal{S} + \tilde{\beta}^{t_1 - t_0} u(q_{n^2}) + \tilde{\beta}^{t_1 - t_0 + 1} V(\hat{n}_1, 0)$$

where $\mathcal{S}$ is the discounted payoff of using the first and second account between t_0 and

$t_1 - 1$.[28] In t_0 there is another feasible strategy: alternatively, the agent could default on the second account with reputation zero instead of trading with the account. This alternative value is then given by:

$$V^* = \mathcal{S}^* + \tilde{\beta}^{t_1-t_0} u(q_{n^2}) + \tilde{\beta}^{t_1-t_0+1} V(\hat{n}_1, 0)$$

where $\mathcal{S}^*$ is the discounted payoff of defaulting on the second account between t_1 and t_0.[29] Since $q_n = q_{n^1} = q_{n^2}$ for all $n < n^1$ and $n < n^2$, this implies $\mathcal{S}^* > \mathcal{S}$ if $q_{n^1} = q_{n^2} > 0$. This is a contradiction.

(b) The borrower does not default before either $n^1 = \hat{n}$ or $n^2 = \hat{n}$ is reached in period t_1. Without loss of generality, let us suppose $n^1 = \hat{n}$. It is easy to verify that we can now apply exactly the same argument as in case 1 and conclude that the borrower has no incentive to increase the reputation on the account with lower levels of reputation.[30] If the agent will not use the second account to accumulate further reputation then he either defaults on the account or he will never use it again:

i. The borrower defaults on the second account. However, according to the same argument as in subcase a), there is a better strategy where the borrower always defaults on the second account instead of accumulating reputation. This is a contradiction.

ii. The borrower never uses the second account again. The deviation value is then given by:

$$\hat{V} = \mathcal{S} + \tilde{\beta}^{t_1-t_0} V(\hat{n}_1, \hat{n}_2)$$

where $\mathcal{S} = [u(q_0) - c(q_0)]\frac{1-\tilde{\beta}^{t_1-t_0}}{1-\tilde{\beta}}$ is the discounted payoff of using the second account between t_1 and t_0. In t_0 there is another feasible strategy: instead of using the second account, the borrower could use only the first account. Let us denote the time when the borrower achieves $n^1 = \hat{n}$ as $t_1^* < t_1$. The alternative value is then given by:

$$V^* = \mathcal{S}^* + \tilde{\beta}^{t_1^*-t_0} V(\hat{n}_1, 0)$$

where $\mathcal{S}^* = [u(q_0) - c(q_0)]\frac{1-\tilde{\beta}^{t_1^*-t_0}}{1-\tilde{\beta}}$ is the discounted payoff of using the first account between t_1^* and t_0. The deviation value can only be strictly beneficial if

[28] The explicit expression is given by $\mathcal{S} = \sum_{n=0}^{\hat{n}_1 - n^1 + n^2} \tilde{\beta}^{\hat{t}_n - t_0}[u(q_0) - c(q_0)]$ where $\hat{t}_n$ is the time where the borrower trades with the second account with reputation n.

[29] The explicit expression is given by $\mathcal{S}^* = \sum_{n=0}^{\hat{n}_1 - n^1 + n^2} \tilde{\beta}^{\hat{t}_n - t_0} u(q_0)$ where $\hat{t}_n$ is the time where the borrower trades with the second account with reputation n.

[30] Essentially, the argument in case 1 is independent of the level of reputation on the second account as long as the second account has lower reputation than the first.

it exceeds the alternative value, $V^* \leq \hat{V}$:

$$\frac{u(q_0) - c(q_0)}{1 - \tilde{\beta}} \geq V(\hat{n}_1, 0)$$

where we used the fact that $V(\hat{n}_1, 0) = V(\hat{n}_1, \hat{n}_2)$ since the second account will never be used from time t_1 onwards. This implies that the deviation value is bounded from above. But then there exists a second alternative strategy where the agent always defaults on the first account from t_1 onwards which yields

$$V^{**} = \frac{u(q_0)}{1 - \tilde{\beta}} > \frac{u(q_0) - c(q_0)}{1 - \tilde{\beta}} \geq \hat{V}.$$

This is a contradiction.

We conclude that opening a second account cannot be optimal. □

5.6.2 Proof of proposition 14

Proof. Consider the following two ND-constraints for $n = N$ and $n = N - 1$:

$$B_N = \beta(1 - p)[V^b(N) - V^b(0)] \geq c(q_N),$$

$$B_{N-1} = \beta(1 - p)[V^b(N) - V^b(0)] \geq c(q_{N-1}),$$

which implies that $B_N = B_{N-1}$. Given we study a not-too-tight equilibrium, this implies $b_N = b_{N-1}$ and from (5.8) it follows that $q_N = q_{N-1}$. Therefore, by (5.12) it follows that $V^b(N) = V^b(N - 1)$. Next, consider the following two ND-constraints for $n = N - 1$ and $n = N - 2$:

$$B_{N-1} = \beta(1 - p)[V^b(N) - V^b(0)] \geq c(q_{N-1}),$$

$$B_{N-2} = \beta(1 - p)[V^b(N - 1) - V^b(0)] \geq c(q_{N-1}).$$

But since $V^b(N) = V^b(N - 1)$ we find that $B_{N-1} = B_{N-2}$. Given we study a not-too-tight equilibrium, this implies $b_{N-1} = b_{N-2}$ and from (5.8) it follows that $q_{N-1} = q_{N-2}$. Therefore, by (5.12) it follows that $V^b(N - 1) = V^b(N - 2)$. Applying this argument recursively implies that $q_n = q$ for all n. But then $V(n) = V(0)$ for all n and thus $B_n = 0$ for all n. As a result, $b_n = 0$ for all n and by (5.8) $q_n = 0$ for all n. □

5.6.3 Proof of proposition 16

Proof. Consider the following sequence: $q_N = \hat{q}$ and $q_n = 0$ for all $n < N-1$. The ND-constraints for $n < N-1$ are thus satisfied. The ND-constraint for $n = N-1$ is satisfied if $0 < q_{N-1} < q_N$. q_{N-1} is determined as such that the ND-constraint for $n = N$ is satisfied:

$$\beta(1-p)\Big[V^b(N) - V^b(0)\Big] = c(q_N).$$

To simplify notation, let us define $(1+r) \equiv \frac{1}{\beta(1-p)}$. We use $S_n = 0$ for all $n < N-1$ and insert (5.12) for $n = N$ and $n = N-1$ into (5.6.3):

$$\frac{1+r}{r}\sigma S_N - \left[\left(\frac{\sigma}{r+\sigma}\right)^{N-1}\frac{1+r}{r+\sigma}\sigma S_{N-1} + \left(\frac{\sigma}{r+\sigma}\right)^{N}\frac{1+r}{r}\sigma S_N\right] = (1+r)c(q_N).$$

We can solve for S_{N-1}:

$$S_{N-1} = \left[\left(\frac{r+\sigma}{\sigma}\right)^{N} - 1\right]\frac{\sigma}{r}S_N - \left(\frac{r+\sigma}{\sigma}\right)^{N} c(q_N). \tag{5.32}$$

Next, we derive conditions under which $S_{N-1} < S_N$ and therefore $q_{N-1} < q_N$. We insert (5.32) into $S_{N-1} < S_N$:

$$S_N > \left[\left(\frac{r+\sigma}{\sigma}\right)^{N} - 1\right]\frac{\sigma}{r}S_N - \left(\frac{r+\sigma}{\sigma}\right)^{N} c(q_N).$$

We can apply the natural logarithm and solve for N:

$$\bar{N} \equiv \frac{\ln\left(\frac{(1+\frac{\sigma}{r})S_N}{\frac{\sigma}{r}S_N - c(q_N)}\right)}{\ln\left(\frac{r+\sigma}{\sigma}\right)} > N.$$

Next, we derive conditions under which $S_{N-1} \geq 0$ and therefore $q_{N-1} \geq 0$. We insert (5.32) into $S_{N-1} \geq 0$:

$$\left[\left(\frac{r+\sigma}{\sigma}\right)^{N} - 1\right]\frac{\sigma}{r}S_N - \left(\frac{r+\sigma}{\sigma}\right)^{N} c(q_N) \geq 0.$$

We can apply the natural logarithm and solve for N:

$$N \geq \frac{\ln\left(\frac{\frac{\sigma}{r}S_N}{\frac{\sigma}{r}S_N - c(q_N)}\right)}{\ln\left(\frac{r+\sigma}{\sigma}\right)} \equiv \underline{N}.$$

One can show that $\bar{N} - \underline{N} = 1$. Given that N is an integer value this means there always exists a unique $\bar{N} > N \geq \underline{N}$ as long as $\frac{\sigma}{r} S_N - c(q_N) > 0$. We can rearrange this condition, use $q_N = \hat{q}$ and find $(1+r)\Big(\frac{\sigma}{r}[u(\hat{q}) - c(\hat{q})] - c(\hat{q})\Big) = \bar{V}(\hat{q})$. But we know that at $q_N = \hat{q}$ the function $\bar{V}(q_N)$ is maximised. Hence, as long as $\bar{V}(q_N) > 0$ for some q_N, it follows that $\bar{V}(\hat{q}) > 0$. To show this, let us define:

$$\mathcal{O}(q) \equiv \frac{\sigma}{r}\Big[u(q) - c(q)\Big] - c(q).$$

Observe that $\mathcal{O}(0) = 0$ and $\lim_{q \to 0} \mathcal{O}'(q) > 0$ implies that $\bar{V}(q_N) > 0$ for some q_N. Therefore, the optimal reputation equilibrium always exists. □

5.6.4 Proof of proposition 18

Proof. We can insert the first constraint and write the Lagrangian:

$$\begin{aligned}
\mathcal{L} = \max_{q,\kappa,\mu^l,\mu^u} & -\kappa(1-p)\left[1 - \frac{\beta p \epsilon}{1-\beta(1-p)}\right] + \beta(1-p)\frac{\sigma(u(q)-c(q))}{1-\beta(1-p)} \\
& - \bar{\mu}\left[\kappa - \frac{\beta\sigma(u(q)-c(q))}{1-\beta(1-p(1-\epsilon))}\right] \\
& + \underline{\mu}\left[\kappa - \frac{c(q)}{1-p(1-\epsilon)}\right]
\end{aligned}$$

where $\underline{\mu}$ and $\bar{\mu}$ are the Lagrangian multipliers associated with each constraint. The sufficient conditions for an optimum are given by:

$$-(1-p)\left[1 - \frac{\beta p \epsilon}{1-\beta(1-p)}\right] + \underline{\mu} - \bar{\mu} = 0, \tag{5.33}$$

$$\beta(1-p)\frac{\sigma(u'(q)-c'(q))}{1-\beta(1-p)} + \bar{\mu}\left[\frac{\beta\sigma(u'(q)-c'(q))}{1-\beta(1-p(1-\epsilon))}\right] - \underline{\mu}\left[\frac{c'(q)}{1-p(1-\epsilon)}\right] = 0, \tag{5.34}$$

$$\frac{\beta\sigma(u(q)-c(q))}{1-\beta(1-p(1-\epsilon))} \geq \kappa, \tag{5.35}$$

$$\kappa \geq \frac{c(q)}{1-p(1-\epsilon)}, \tag{5.36}$$

$$\bar{\mu}\left[\kappa - \frac{\beta\sigma(u(q)-c(q))}{1-\beta(1-p(1-\epsilon))}\right] = 0,$$

$$\underline{\mu}\left[\kappa - \frac{c(q)}{1-p(1-\epsilon)}\right] = 0.$$

First, we want to argue that $\underline{\mu} > 0$. Suppose not. Then $\underline{\mu} = 0$. By (5.33)

$$-(1-p)\left[1 - \frac{\beta p \epsilon}{1-\beta(1-p)}\right] - \bar{\mu} = 0.$$

But this is a contradiction since $\left[1 - \frac{\beta p \epsilon}{1-\beta(1-p)}\right] > 0$ and $\bar{\mu} \geq 0$. Thus, $\underline{\mu} > 0$ and

$$\kappa = \frac{c(q)}{1-p(1-\epsilon)}.$$

Next, suppose that $\bar{\mu} = 0$. In that case, we solve (5.33) for $\underline{\mu}$ and insert into (5.34):

$$\frac{u'(q)}{c'(q)} = 1 + \frac{1-\beta(1-p(1-\epsilon))}{\sigma\beta}.$$

Because the left-hand side spans every positive real number and is strictly decreasing in q while the right-hand-side is a positive real number, it follows that a unique q solves that equation. Let us denote this candidate solution by q_1. For this to be a solution, q_1 must satisfy (5.35). To see that (5.35) is indeed satisfied at q_1, suppose that $\bar{\mu} > 0$. In that case (5.35) and (5.36) imply:

$$\frac{c(q)}{1-p(1-\epsilon)} = \frac{\beta\sigma(u(q)-c(q))}{1-\beta(1-p(1-\epsilon))}.$$

This equation has to be satisfied for some q. To study the implied q, let us define:

$$\mathcal{A}(q) \equiv \frac{\beta\sigma(u(q)-c(q))}{1-\beta(1-p(1-\epsilon))} - \frac{c(q)}{1-p(1-\epsilon)}. \tag{5.37}$$

One can easily see that $\mathcal{A}(0) = 0$, $\lim_{q\to 0} \mathcal{A}'(q) = \infty$ and $\mathcal{A}(\bar{q}) < 0$. This has a couple of implications. First, there must be two solutions to (5.37): $q = 0$ and some $\dot{q} > 0$. Second, it follows that $0 < \tilde{q} < \dot{q}$ where $\tilde{q} = \arg\max_q \mathcal{A}(q)$. Third, as we assume full redistribution of the fees, $\tilde{q}$ is uniquely pinned down by:

$$\frac{u'(\tilde{q})}{c'(\tilde{q})} = 1 + \frac{1-\beta}{\sigma\beta}.$$

But observe that $\hat{q} = \tilde{q}$. But then we know that $\hat{q} < \dot{q}$ and therefore $\bar{\mu} > 0$. Hence, the optimal allocation is $q = \hat{q}$. $\square$

Bibliography

Acharya, S. and K. Dogra (2022). The side effects of safe asset creation. *Journal of the European Economic Association 20*(2), 581–625.

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

Ahnert, T. and M. Macchiavelli (2021). Safe assets and financial fragility: Theory and evidence. *FRB of New York Staff Reports*.

Aiyagari, S. R. and M. Gertler (1991). Asset returns with transactions costs and uninsured individual risk. *Journal of Monetary Economics 27*(3), 311–331.

Alon, T., E. Swanson, et al. (2011). Operation twist and the effect of large-scale asset purchases. *FRB of San Francisco Economic Letter 13*, 1–5.

Alvarez, F. and U. J. Jermann (2000). Efficiency, equilibrium, and asset pricing with risk of default. *Econometrica 68*(4), 775–797.

Amihud, Y. and H. Mendelson (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics 17*(2), 223–249.

Amihud, Y., H. Mendelson, and L. H. Pedersen (2006). Liquidity and asset prices. *Foundations and Trends in Finance 1*(4), 269–364.

Anbil, S., A. Anderson, and Z. Senyuz (2020). What happened in money markets in September 2019? *FEDS Notes*.

Andres, J., J. D. López-Salido, and E. Nelson (2004). Tobin's imperfect asset substitution in optimizing general equilibrium. *Journal of Money, Credit and Banking*, 665–690.

Araujo, L. (2004). Social norms and money. *Journal of Monetary Economics 51*(2), 241–256.

Azzimonti, M. and P. Yared (2019). The optimal public and private provision of safe assets. *Journal of Monetary Economics 102*, 126–144.

Bacchetta, P. and E. van Wincoop (2021). Puzzling exchange rate dynamics and delayed portfolio adjustment. *Journal of International Economics 131.*

Baklanova, V., A. Copeland, and R. McCaughrin (2015). Reference guide to U.S. repo and securities lending markets. *FRB of New York Staff Reports* (740).

Barclay, M. J., T. Hendershott, and K. Kotz (2006). Automation versus intermediation: Evidence from Treasuries going off the run. *Journal of Finance 61*(5), 2395–2414.

Barro, R. J., J. Fernández-Villaverde, O. Levintal, and A. Mollerus (2022). Safe assets. *The Economic Journal 132*(646), 2075–2100.

Bech, M., A. Illes, U. Lewrick, and A. Schrimpf (2016). Hanging up the phone – electronic trading in fixed income markets and its implications. *BIS Quarterly Review*.

Bessembinder, H., C. Spatt, and K. Venkataraman (2020). A survey of the microstructure of fixed-income markets. *Journal of Financial and Quantitative Analysis 55*(1), 1–45.

Bethune, Z., T.-W. Hu, and G. Rocheteau (2018). Indeterminacy in credit economies. *Journal of Economic Theory 175*, 556–584.

Bewley, T. (1979). The optimum quantity of money. Technical report, Discussion Paper.

Bowman, D., J. Louria, M. McCormick, and M.-F. Styczynski (2017). The cleared bilateral repo market and proposed repo benchmark rates. *FEDS Notes*.

Brown, M. and C. Zehnder (2007). Credit reporting, relationship banking, and loan repayment. *Journal of Money, Credit and Banking 39*(8), 1883–1918.

Brunner, K. and A. H. Meltzer (1972). Money, debt, and economic activity. *Journal of Political Economy 80*(5), 951–977.

Brunnermeier, M. and V. Haddad (2014). Safe assets.

Brunnermeier, M. K., S. Merkel, and Y. Sannikov (2022). Safe assets: A dynamic retrading perspective. Technical report, mimeo.

Buraschi, A. and A. Jiltsov (2007). Habit formation and macroeconomic models of the term structure of interest rates. *The Journal of Finance 62*(6), 3009–3063.

Caballero, R. J. and E. Farhi (2013). A model of the safe asset mechanism (SAM): Safety traps and economic policy. Technical report, National Bureau of Economic Research.

Caballero, R. J., E. Farhi, and P.-O. Gourinchas (2016). Safe asset scarcity and aggregate demand. *American Economic Review 106*(5), 513–518.

Caballero, R. J., E. Farhi, and P.-O. Gourinchas (2017). The safe assets shortage conundrum. *Journal of Economic Perspectives 31*(3), 29–46.

Campbell, J. Y. and L. M. Viceira (2002). *Strategic asset allocation: portfolio choice for long-term investors.* Clarendon Lectures in Economics.

Carapella, F. and S. Williamson (2015). Credit markets, limited commitment, and government debt. *The Review of Economic Studies 82*(3), 963–990.

Cashin, D. B., E. Syron Ferris, E. Klee, and C. Stevens (2017). Take it to the limit: The debt ceiling and Treasury yields.

Chambers, D. and E. Dimson (2009). IPO underpricing over the very long run. *The Journal of Finance 64*(3), 1407–1443.

Cho, I.-K. and D. M. Kreps (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics 102*(2), 179–221.

Christensen, J. H., N. Mirkov, et al. (2021). The safety premium of safe assets. Technical report, Swiss National Bank.

Christensen, J. H. E., J. A. Lopez, and P. Shultz (2017). Do all new Treasuries trade at a premium? *FRB of San Francisco Economic Letter* (3).

Christensen, J. H. E., J. A. Lopez, and P. J. Shultz (2020). Is there an on-the-run premium in TIPS? *The Quarterly Journal of Finance 10*(2).

Constantinides, G. M. (1986). Capital market equilibrium with transaction costs. *Journal of Political Economy 94*(4), 842–862.

Constantinides, G. M. (2002). Rational asset prices. *The Journal of Finance 57*(4), 1567–1591.

Constantinides, G. M. and D. Duffie (1996). Asset pricing with heterogeneous consumers. *Journal of Political Economy 104*(2), 219–240.

Corradin, S. and A. Maddaloni (2020). The importance of being special: Repo markets during the crisis. *Journal of Financial Economics 137*(2), 392–429.

D'Amico, S. and N. A. Pancost (2022). Special repo rates and the cross-section of bond prices: The role of the special collateral risk premium. *Review of Finance 26*(1), 117–162.

Dang, T. V., G. Gorton, and B. Holmstrom (2009). Opacity and the optimality of debt for liquidity provision. *Manuscript Yale University*.

Dang, T. V., G. Gorton, and B. Holmström (2010). Financial crises and the optimality of debt for liquidity provision. *Unpublished working paper. Yale School of Management*.

Davis, M. H. and A. R. Norman (1990). Portfolio selection with transaction costs. *Mathematics of Operations Research 15*(4), 676–713.

DeMarzo, P. and D. Duffie (1999). A liquidity-based model of security design. *Econometrica 67*(1), 65–99.

Di Tella, S. (2020). Risk premia and the real effects of money. *American Economic Review 110*(7), 1995–2040.

Diamond, W. and P. Van Tassel (2021). Risk-free rates and convenience yields around the world. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Doerr, S., E. Eren, and S. Malamud (2023). Money market funds and the pricing of near-money assets. *BIS Working Paper*.

Du, W., B. Hèbert, and W. Li (2023). Intermediary balance sheets and the Treasury yield curve. *Journal of Financial Economics 150*(3).

Duffie, D. (1996). Special repo rates. *The Journal of Finance 51*(2), 493–526.

Duffie, D. (2020). Still the worlds safe haven. *Redesigning the US Treasury market after the COVID-19 crisis. Hutchins Center on Fiscal and Monetary Policy at Brookings. Available online: https://www. brookings. edu/research/still-the-worlds-safe-haven (accessed on 28 November 2021)*.

Duffie, D. (2023). Resilience redux in the US Treasury market. Jackson Hole Symposium.

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

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

Duffie, D., T. F. Geithner, P. Parkinson, and J. Stein (2021). U.S. Treasury markets: steps toward increased resilience. https://group30.org/publications/detail/4950.

Duffie, D. and F. M. Keane (2023). Market-function asset purchases. *FRB of New York Staff Report* (1054).

Duffie, D. and T.-s. Sun (1990). Transactions costs and portfolio choice in a discrete-continuous-time setting. *Journal of Economic Dynamics and Control 14*(1), 35–51.

Dumas, B. and E. Luciano (1991). An exact solution to a dynamic portfolio choice problem under transactions costs. *The Journal of Finance 46*(2), 577–595.

Durham, J. B. and R. Perli (2022). *The Research Handbook of Financial Markets*, Chapter "The Treasury and the When-Issued Market". Forthcoming from Edward Elgar.

Elul, R. and P. Gottardi (2015). Bankruptcy: Is it enough to forgive or must we also forget? *American Economic Journal: Microeconomics 7*(4), 294–338.

Engel, C. and S. P. Y. Wu (2023). Liquidity and exchange rates: An empirical investigation. *The Review of Economic Studies 90*(5), 2395–2438.

Epstein, L. G. and S. E. Zin (1989). Substitution, risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica* (4), 937–969.

Eren, E. and P. Wooldridge (2021). Non-bank financial institutions and the functioning of government bond markets. *BIS Papers* (119).

Filipovic, D. and A. B. Trolle (2013). The term structure of interbank risk. *Journal of Financial Economics 109*(3), 707–733.

Fleckenstein, M. and F. Longstaff (2023). Treasury richness. Technical report, NBER.

Fleming, M. and F. Keane (2021). The netting efficiencies of marketwide central clearing. *FRB of New York Staff Reports* (964).

Fleming, M., G. Nguyen, and J. Rosenberg (2022). How do Treasury dealers manage their positions? *FRB of New York Staff Reports* (299).

Fleming, M. J. and K. D. Garbade (2002). When the back office moved to the front burner: Settlement fails in the Treasury market after 9/11. *FRB of New York Economic Policy Review*.

Fleming, M. J. and K. D. Garbade (2004). Repurchase agreements with negative interest rates. *Current Issues in Economics and Finance 10*(5).

Fleming, M. J. and K. D. Garbade (2005). Explaining settlement fails. *Current Issues in Economics and Finance 11*(9).

Fleming, M. J., F. M. Keane, A. Martin, and M. McMorrow (2014). What explains the June spike in Treasury settlement fails? Liberty Street Economics.

Fleming, M. J., B. Mizrach, and G. Nguyen (2018). The microstructure of a U.S. Treasury ECN: The BrokerTec platform. *Journal of Financial Markets 40*, 2–22.

Fleming, M. J., G. Nguyen, and J. V. Rosenberg (2024). How do treasury dealers manage their positions? *FRB of New York Staff Report* (299).

Friedman, E. J. and P. Resnick (2004). The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy 10*(2), 173–199.

Frost, J., L. Logan, A. Martin, P. McCabe, F. Natalucci, and J. Remache (2015). Overnight RRP operations as a monetary policy tool: Some design considerations. *Finance and Economics Discussion Series* (2015-010).

Gabaix, X. (2012). Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *The Quarterly Journal of Economics 127*(2), 645–700.

Gagnon, J., M. Raskin, J. Remache, and B. P. Sack (2011). Large-scale asset purchases by the Federal Reserve: did they work? *Economic Policy Review 17*(1), 41.

Garbade, K. D., F. M. Keane, L. Logan, A. Stokes, and J. Wolgemuth (2010). The introduction of the TMPG Fails Charge for U.S. Treasury securities. *FRB of New York Economic Policy Review*.

Gârleanu, N. and L. H. Pedersen (2013). Dynamic trading with predictable returns and transaction costs. *The Journal of Finance 68*(6), 2309–2340.

Goldreich, D. (2007). Underpricing in discriminatory and uniform-price Treasury auctions. *Journal of Financial and Quantiative Analysis 42*(2), 443–466.

Goldreich, D., B. Hanke, and P. Nath (2005). The price of future liquidity: time-varying liquidity in the U.S. Treasury market. *Review of Finance 9*(1), 1–32.

Golec, P. and E. Perotti (2017). Safe assets: a review.

Gorton, G. (2017). The history and economics of safe assets. *Annual Review of Economics 9*, 547–586.

Gorton, G. (2020). The regulation of private money. *Journal of Money, Credit and Banking 52*(S1), 21–42.

Gorton, G., S. Lewellen, and A. Metrick (2012). The safe-asset share. *American Economic Review 102*(3), 101–106.

Gorton, G. and G. Ordonez (2022). The supply and demand for safe assets. *Journal of Monetary Economics 125*, 132–147.

Gorton, G. and G. Pennacchi (1990). Financial intermediaries and liquidity creation. *The Journal of Finance 45*(1), 49–71.

Greenwood, R. and D. Vayanos (2010). Price pressure in the government bond market. *American Economic Review 100*(2), 585–590.

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

Grossman, S. J. and G. Laroque (1990). Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods. *Econometrica: Journal of the Econometric Society*, 25–51.

Gu, C., F. Mattesini, C. Monnet, and R. Wright (2013). Endogenous credit cycles. *Journal of Political Economy 121*(5), 940–965.

Gu, C., F. Mattesini, and R. Wright (2016). Money and credit redux. *Econometrica 84*(1), 1–32.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2010). The TIPS yield curve and inflation compensation. *American Economic Journal: Macroeconomics 2*(1), 70–92.

Hamilton, J. D. and J. C. Wu (2012). The effectiveness of alternative monetary policy tools in a zero lower bound environment. *Journal of Money, Credit and Banking 44*, 3–46.

He, Z., A. Krishnamurthy, and K. Milbradt (2019). A model of safe asset determination. *American Economic Review 109*(4), 1230–1262.

He, Z., S. Nagel, and Z. Song (2022). Treasury inconvenience yields during the COVID-19 crisis. *Journal of Financial Economics 143*(1), 57–79.

Heaton, J. and D. Lucas (2000). Portfolio choice in the presence of background risk. *The Economic Journal 110*(460), 1–26.

Heaton, J. and D. J. Lucas (1996). Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy 104*(3), 443–487.

Holmström, B. and J. Tirole (2001). LAPM: A liquidity-based asset pricing model. *The Journal of Finance 56*(5), 1837–1867.

Hua, X. and J. Watson (2022). Starting small in project choice: A discrete-time setting with a continuum of types. *Journal of Economic Theory 204*, 105490.

Huang, M. (2003). Liquidity shocks and equilibrium liquidity premia. *Journal of Economic Theory 109*(1), 104–129.

Huh, Y. and S. Infante (2021). Bond market intermediation and the role of repo. *Journal of Banking and Finance 122*.

Iachan, F. S., P. T. Nenov, and A. Simsek (2021). The choice channel of financial innovation. *American Economic Journal: Macroeconomics 13*(2), 333–372.

Jiang, Z., A. Krishnamurthy, and H. Lustig (2021). Foreign safe asset demand and the dollar exchange rate. *The Journal of Finance 76*(3), 1049–1089.

Kacperczyk, M., C. Perignon, and G. Vuillemey (2021). The private production of safe assets. *The Journal of Finance 76*(2), 495–535.

Kehoe, T. J. and D. K. Levine (1993). Debt-constrained asset markets. *The Review of Economic Studies 60*(4), 865–888.

Klingler, S. and S. Sundaresan (2023). Diminishing Treasury convenience premiums: Effects of dealers' excess demand and balance sheet constraints. *Journal of Monetary Economics 135*, 55–69.

Kocherlakota, N. and N. Wallace (1998). Incomplete record-keeping and optimal payment arrangements. *Journal of Economic Theory 81*(2), 272–289.

Krishnamurthy, A. (2002). The bond/old-bond spread. *Journal of Financial Economics 66*(2-3), 463–506.

Krishnamurthy, A. and W. Li (2023). The demand for money, near-money, and Treasury bonds. *The Review of Financial Studies 36*(5), 2091–2130.

Krishnamurthy, A. and A. Vissing-Jorgensen (2011). The effects of quantitative easing on interest rates: Channels and implications for policy. *Brookings Papers on Economic Activity 43*(2), 215–287.

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

Krishnamurthy, A. and A. Vissing-Jorgensen (2015). The impact of Treasury supply on financial sector lending and stability. *Journal of Financial Economics 118*(3), 571–600.

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

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

Lagos, R. and R. Wright (2005). A unified framework for monetary theory and policy analysis. *Journal of Political Economy 113*(3), 463–484.

Lettau, M. and J. A. Wachter (2011). The term structures of equity and interest rates. *Journal of Financial Economics 101*(1), 90–113.

Li, D. and N. Schürhoff (2018). Dealer networks. *The Journal of Finance 74*(1), 91–144.

Liu, S. and C. Wu (2017). Repo counterparty risk and on-/off-the-run Treasury spreads. *Review of Asset Pricing Studies 7*(1), 81–143.

Lo, A. W., H. Mamaysky, and J. Wang (2004). Asset prices and trading volume under fixed transactions costs. *Journal of Political Economy 112*(5), 1054–1090.

Lotz, S. and C. Zhang (2015). Money and credit as means of payment: A new monetarist approach. *Journal of Economic Theory 164*, 68–100.

Macchiavelli, M. and L. Pettit (2021). Liquidity regulation and financial intermediaries. *Journal of Financial and Quantitative Analysis 56*(6), 2237 – 2271.

Mallick, S., M. S. Mohanty, and F. Zampolli (2017). Market volatility, monetary policy and the term premium.

Marte, J. (2021). NY Fed eases eligibility requirements for reverse repo facility. Reuters article.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995). *Microeconomic Theory.* Oxford University Press, Inc.

Modigliani, F. and R. Sutch (1966). Innovations in interest rate policy. *The American Economic Review 56*(1/2), 178–197.

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

Pasquariello, P. and C. Vega (2009). The on-the-run liquidity phenomenon. *Journal of Financial Economics 92*(1), 1–24.

Qin, K., L. Zhou, Y. Afonin, L. Lazzaretti, and A. Gervais (2021). CeFi vs. DeFi–comparing centralized to decentralized finance. *arXiv preprint arXiv:2106.08157*.

Schrimpf, A., H. S. Shin, and V. Sushko (2020). Leverage and margin spirals in fixed income markets during the COVID-19 crisis. *BIS Bulletin* (2).

Spindt, P. A. and R. W. Stolz (1992). Are US Treasury bills underpriced in the primary market? *Journal of Banking and Finance 16*(5), 891–908.

Strebulaev, I. (2002). Liquidity and asset pricing: Evidence from the U.S. Treasury securities market. Unpublished working paper.

Sunderam, A. (2015). Money creation and the shadow banking system. *The Review of Financial Studies 28*(4), 939–977.

Swanson, E. T. (2011). Let's twist again: a high-frequency event-study analysis of operation twist and its implications for QE2. *Brookings Papers on Economic Activity 2011*(1), 151–188.

Tobin, J. (1958). Liquidity preference as behavior towards risk. *The Review of Economic Studies 25*(2), 65–86.

Tobin, J. (1961). Money, capital, and other stores of value. *The American Economic Review 51*(2), 26–37.

Tobin, J. (1969). A general equilibrium approach to monetary theory. *Journal of Money, Credit and Banking 1*(1), 15–29.

U.S. Department of the Treasury (2022). Additional public transparency in Treasury markets.

U.S. Department of the Treasury, Board of Governors of the Federal Reserve System, Federal Reserve Bank of New York, U.S. Securities and Exchange Commission, and U.S. Commodity Futures Trading Commission (2015). Joint staff report: The U.S. Treasury market on October 15, 2014.

Vayanos, D. (1998). Transaction costs and asset prices: A dynamic equilibrium model. *The Review of Financial Studies 11*(1), 1–58.

Vayanos, D. (2004). Flight to quality, flight to liquidity, and the pricing of risk.

Vayanos, D. and J.-L. Vila (1999). Equilibrium interest rate and liquidity premium with transaction costs. *Economic Theory 13*, 509–539.

Vayanos, D. and J.-L. Vila (2021). A preferred-habitat model of the term structure of interest rates. *Econometrica 89*(1), 77–112.

Vayanos, D. and P.-O. Weill (2008). A search-based theory of the on-the-run phenomenon. *The Journal of Finance 63*(3), 1361–1398.

Vissing-Jorgensen, A. (2002). Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy 110*(4), 825–853.

Vissing-Jorgensen, A. (2021). The Treasury market in spring 2020 and the response of the Federal Reserve. *NBER Working Paper Series* (29128).

Wachter, J. A. (2006). A consumption-based model of the term structure of interest rates. *Journal of Financial Economics 79*(2), 365–399.

Wang, C.-C. and Y. Li (2023). Anonymous credit. *MPRA Paper* (118480).

Watson, J. (1999). Starting small and renegotiation. *Journal of Economic Theory 85*(1), 52–90.

Wells, M. (2023). Averting a Treasury market crisis. *Econ Focus*, 14–17.

Xiong, W. and H. Yan (2010). Heterogeneous expectations and bond markets. *The Review of Financial Studies 23*(4), 1433–1466.

Statement of authorship

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Koautorenschaften sowie alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe o des Gesetzes vom 5. September 1996 über die Universität zum Entzug des aufgrund dieser Arbeit verliehenen Titels berechtigt ist.

Own English translation: I declare herewith that I wrote this thesis on my own and used no other than the listed referenced sources. I have marked all co-authorships and all passages taken word for word or in meaning from sources as such. I understand that if I do not follow these rules that the Senate of the University of Bern is authorised to revoke the title awarded on the basis of this thesis according to Article 36, paragraph 1, letter o of the University Act of September 5th, 1996.

Bern, 20 June 2024

Fabienne Schneider