What's So Inconvenient About TIPS?

Athanasios Geromichalos – University of California, Davis

Lucas Herrenbrueck – Simon Fraser University

Changhyun Lee – University of California, Davis

Sukjoon Lee – New York University Shanghai

First version: October 2024
This version: March 2025

ABSTRACT

We build on recent developments in the theory of money and liquidity to provide a qualitative and quantitative explanation for the well-known TIPS illiquidity vis-à-vis non-inflation-protected Treasuries. Our model does not assume exogenous differences between the markets where the two assets are traded or the investors who hold them; instead, an asset's liquidity is endogenous and depends on the trading and market entry decisions of investors. The model offers a powerful amplification mechanism that operates upon only one difference between TIPS and Treasuries: the far greater supply of the latter. We also quantify how much the Treasury leaves on the table by issuing securities that are not as highly valued by the market as nominal Treasuries, thus informing the recent debate. The model does not necessarily imply that the Treasury should phase out TIPS, as some have suggested, only that it should seek to reduce segmentation and ensure that TIPS trade alongside nominal Treasuries in *consolidated* secondary markets.

JEL Classification: E31, E43, E52, G12

Keywords: Monetary-search models, OTC markets, Endogenous liquidity, Treasury securities, Treasury Inflation-Protected Securities (TIPS)

Contact: ageromich@ucdavis.edu, herrenbrueck@sfu.ca, chzlee@ucdavis.edu, sukjoon.lee@nyu.edu

We would like to thank Andrew Atkenson, Jeremie Banet, Nicolas Caramp, Yuriy Gorodnichenko, and Guillaume Rocheteau for useful comments and suggestions.

"So my classmate, Larry Summers, introduced TIPS, Treasury Inflation Protected Securities, and he, you know, announced this was a fantastic idea and was going to change radically the whole markets. And then nobody traded them, and they offered astronomical interest rates, real interest rates, to get anyone to buy them. And so they were nicknamed—it's really bad on camera—but their nickname became Totally Illiquid Pieces of ..."

— John Geanakoplos

1 Introduction

A large body of research reveals that Treasury Inflation-Protected Securities (TIPS) are characterized by lower trading volume, longer turnaround times, and greater bid-ask spreads than those observed for (non-inflation-indexed) U.S. Treasuries—in other words, they are less liquid—and that these differences are priced.[1] Of course, it is not uncommon for two asset classes of otherwise similar characteristics to be traded with different degrees of liquidity; for example, Treasuries sell at higher prices than corporate or municipal bonds of similar characteristics, a difference commonly attributed to the convenience yield that investors are willing to pay for securities issued by the U.S. Treasury (Krishnamurthy and Vissing-Jorgensen, 2012). But TIPS are issued by the same agency as regular Treasuries. If the "convenience yield for Treasury-issued bonds" narrative does not apply, how can one explain the documented illiquidity of TIPS vis-à-vis their non-inflation-protected counterparts?

In this study, we use recent developments in the theory of money and liquidity to rationalize this puzzling observation, both qualitatively and quantitatively. Our model is a monetary dynamic general-equilibrium model with two main ingredients. The first is an empirically relevant concept of asset liquidity. Agents receive an idiosyncratic shock that determines whether they will be active consumers that period; since money is the unique medium of exchange, active consumers visit secondary asset markets to boost their liquidity (i.e., to sell assets for money), and inactive consumers are the providers of that liquidity. The secondary asset markets are modeled as Over-the-Counter (OTC) markets characterized by search and bargaining. The second ingredient is a market entry decision by the agents; there are two assets (A and B), each asset trades in a distinct secondary market, and agents choose to visit the market where they expect to find the best terms. Crucially, the markets are *ex-ante* identical; thus, any liquidity differences are

[1] See Fleming and Krishnan (2012) and Fleckenstein, Longstaff, and Lustig (2014) among many others.

not hard-wired but are the outcome of agents' trading and entry decisions.

To sell an asset in the secondary market, our model requires that the seller own that asset; thus, the market choice of a potential seller is 'locked in' by the type of asset they own. Asset buyers' OTC market entry, on the other hand, is flexible since their money is good to buy any asset. As a result, when market A has any kind of advantage over market B (in a sense to be made precise below), asset buyers rush into market A more eagerly than sellers, which implies a higher trade probability for sellers in that market. This higher sell-probability manifests itself as a greater *ex-ante* willingness to pay higher prices for the asset that agents are expecting to sell more easily 'down the road'. The liquidity advantage is complemented by a general-equilibrium mechanism: understanding that matching in market B will be harder, agents who hold that asset also choose to bring more money, reducing their reliance on an additional liquidity boost and decreasing the potential trade surplus in market B, which reinforces the dominance of market A. When coupled with increasing returns to scale (IRS) in OTC market matching, the two channels just described are further amplified because IRS promote concentration of investors in the market with the original advantage. We find that with even a minimal degree of IRS, this amplification mechanism can be so powerful that asset demand curves become upward sloping, that is, assets that come in larger quantities are *endogenously* more liquid.

Given this discussion, and letting asset A stand for Treasuries and asset B for TIPS, we can now explain how the proposed model can rationalize the illiquidity of TIPS. In our framework, Treasuries sell at a higher liquidity premium than TIPS, not because of some exogenous advantage in the provision of liquidity (or even provision of utility), but due to the sheer size of their supply. The larger supply of Treasuries implies that more of them are likely to be offered for sale in the secondary market, thus, potential meetings in the secondary market for Treasuries are likely to generate a larger trading surplus. *This* is the advantage that originally attracts more buyers into the market for Treasuries. The expected presence of more buyers in that market makes it more advantageous to hold Treasuries in the first place; importantly, holding Treasuries today entails seeking to sell Treasuries tomorrow, if the holder is hit by the liquidity shock. With IRS, increased entry into the market for Treasuries implies higher trading probabilities for both buyers and sellers; but of these, it is the higher sell-probability, or, as financial economists would say, the lower *liquidity risk* (the risk of trying to sell and being unable to do so), that induces agents to pay a higher liquidity premium for Treasuries.

The discussion so far shows how our theory can explain TIPS illiquidity in a qualitative fashion. We also demonstrate that a calibrated version of our model can quantitatively match the yields of Treasuries and TIPS, as well as their differentials observed in the data. More precisely, we calibrate our model to key facts about the U.S. Treasury market, including yields and liquidity premia, and show that our theory can quantitatively account for the illiquidity of TIPS, as long as the bargaining power of sellers in OTC markets is relatively high. This result is intuitive: in our model, agents are willing to pay liquidity premia for assets because they predict that they will be able to *sell* them for money in the OTC round of trade. A high bargaining power for sellers in that market guarantees that this crucial channel will be quantitatively active. Notably, the model can perfectly match the data with a minimal degree of IRS in OTC matching, and the degree of IRS needed to match the data is decreasing in the seller's bargaining power.[2]

Next, we use our model to perform a policy-relevant counterfactual exercise. Prior to the introduction of TIPS in 1997, several reports described how fantastic this new product would be, calculating the savings the Treasury would achieve by issuing TIPS. (See the epigraph and Section 5 for details.) With hindsight, the introduction of TIPS did not save taxpayers billions, as had been predicted, but in fact led to the opposite outcome because TIPS turned out to be less liquid than Treasuries. This spurred a debate over whether TIPS had been a mistake, with some economists wondering how much money the Treasury is leaving on the table by issuing TIPS. Our model offers an opportunity for a counterfactual analysis that can *put a number* on the cost of issuing TIPS, thus informing the recent debate. Through the lens of our model, TIPS are less liquid than Treasuries because they trade in segmented secondary markets and come in lower supply. Thus, our model implies that the Treasury does not need to phase out TIPS, as has been suggested. An alternative approach would seek to reduce secondary market segmentation and ensure that TIPS trade alongside nominal Treasuries in consolidated secondary markets. We find that if TIPS and Treasuries traded in a perfectly consolidated secondary market, the Treasury would save up to 700 million dollars per year on the issuance of 10-year Treasuries and TIPS during the post-Great Recession to pre-Covid period. We also provide a discussion of how such secondary market consolidation could be achieved in reality.

Even though our model does not *need* to assume exogenous differences between the markets where the two asset types trade or the investors who hold them, such differences may indeed be relevant in real-world Treasury markets. Thus, following some of the related finance literature, we also explore (in Section 6) an extension of the model where a

---

[2] Our OTC matching function features a key parameter, $\rho \in [0, 1]$, which measures the intensity of returns to scale in matching: $\rho = 0$ captures constant returns to scale (CRS), and, on the other extreme, $\rho = 1$ captures congestion-free matching. As long as the seller's bargaining power is higher than 0.7, our calibrated model can perfectly match the data on Treasury and TIPS yields (and their differential) with a degree of IRS that is below 0.01, i.e., in the *neighborhood of CRS*. Note that Duffie, Gârleanu, and Pedersen (2005) and most of the papers that adopt their framework assume a congestion-free matching function, i.e., one with $\rho = 1$.

fraction of agents "buy and hold" TIPS (see, for example, Andreasen, Christensen, and Riddell, 2021). The goal of this exercise is to study how this new (and realistic) ingredient interacts with our amplification mechanism in general equilibrium. The presence of buy *and hold* investors effectively reduces the supply of TIPS that will be available for trade in the respective OTC market. This reduction is anticipated by regular (non-buy-and-hold) agents, and through the endogenous market entry mechanism of our model, it exacerbates the illiquidity of TIPS. In short, the presence of the buy-and-hold investors further *enhances* our quantitative results, and this is manifested in two ways. First, the degree of IRS in matching needed to fit the data drops even further in the model with buy-and-hold investors; second, for given parameter values, the gain the U.S. Treasury would obtain if the two secondary markets were perfectly consolidated increases even further. (Under certain parametrizations, the savings the Treasury would achieve are 50% greater than the baseline specification; see Table 5 for details.)

A remarkable finding related to this exercise is that even in the presence of buy-and-hold investors (and even if their measure is large), the liquidity premia of Treasuries and TIPS are equalized, if these two asset types trade in a unique, consolidated secondary market. Simply put, the addition of buy-and-hold investors in the model (an ingredient which is admittedly realistic) cannot explain the illiquidity of TIPS in its own right, and assuming some segmentation between the secondary markets for TIPS and Treasuries is a *necessary* condition for capturing the well-documented illiquidity of TIPS. At the same time, the addition of buy-and-hold investors makes the message of our paper even stronger, as it highlights that the U.S. Treasury leaves substantial profits on the table by having Treasuries and TIPS trade in segmented secondary markets.

The illiquidity of TIPS has been documented by numerous papers in empirical finance, e.g., Sack and Elsasser (2004), Campbell, Shiller, and Viceira (2009), Dudley, Roush, and Ezer (2009), Gürkaynak, Sack, and Wright (2010), Fleming and Krishnan (2012), Fleckenstein et al. (2014), Abrahams, Adrian, Crump, Moench, and Yu (2016), Pflueger and Viceira (2016), D'Amico, Kim, and Wei (2018), and Andreasen et al. (2021). These papers focus on estimating the yield differential between TIPS and Treasuries after controlling for inflation expectations. On the model side, most papers in this literature assume exogenous differences between the markets where these assets are traded and/or the investors who hold them. Our paper uses the theoretical framework of Geromichalos, Herrenbrueck, and Lee (2023a) and shows that the TIPS premium can be rationalized, both qualitatively and quantitatively, based only on a difference that is perfectly observed in the data: the far greater supply of Treasuries. However, as explained, motivated by some papers in this literature, we also explore an extension of our model that imposes exogenous differences on the investors who hold the two types of assets. Specifically, we study a version of the model where a fraction of agents "buy and hold" TIPS, and we show that the presence of such investors further enhances our quantitative results.

Since in our model the liquidation of assets takes place in an OTC secondary asset market, our paper is related to the seminal work of Duffie et al. (2005), which studies how frictions in OTC financial markets affect asset prices. Extensions of this framework have been studied by Weill (2007, 2008), Vayanos and Wang (2007), Vayanos and Weill (2008), Lagos and Rocheteau (2009), Lagos, Rocheteau, and Weill (2011), Afonso and Lagos (2015), Chang and Zhang (2015), and Üslü (2019), among others. In addition to focusing on a different research question, the illiquidity of TIPS, our paper also differs methodologically from this literature, starting with the very concept of liquidity. We employ a *monetary* model, in which agents sell assets (in distinct secondary markets) for cash upon learning of a consumption opportunity, whereas in those papers, agents differ in the utility flow derived from holding an asset. Consequently, in those papers agents can buy assets using transferable utility, while in our model agents who wish to buy assets must incur the cost of carrying money. Importantly, by incorporating both financial market frictions and monetary factors (such as inflation and the nominal policy rate), our approach provides a more comprehensive understanding of the forces that drive regular and inflation-indexed Treasury prices, enhancing the model's quantitative ability to capture empirical patterns in asset yields.[3]

Our paper belongs to a literature arguing that asset liquidity can be employed to resolve long-standing puzzles in asset-pricing theory (Lagos, 2010; Geromichalos and Simonovska, 2014; Geromichalos, Herrenbrueck, and Salyer, 2016; Herrenbrueck and Wang, 2023; Lee and Jung, 2020; Caramp and Singh, 2023). The inclusion of the secondary market, where agents rebalance their portfolios depending on liquidity needs, imbues our assets with "indirect liquidity" as in Herrenbrueck and Geromichalos (2017): assets that can be exchanged easily for money enable an agent to carry less money in the first place, and therefore affect the broad supply of "liquidity" even if they are never used as payment "directly". Examples of papers that belong in the indirect liquidity literature include

[3] Weill (2008) and Vayanos and Weill (2008) also seek to explain price differentials through endogenous liquidity differences, with agents self-sorting into separate markets. However, in these papers, agents can hold only 0 or 1 unit of the asset, thus, an increase in an asset's supply necessarily means that more agents must hold it. Our model is based on the New-Monetarist framework allowing us to consider perfectly divisible asset (and money) holdings, so that agents choose freely how much to hold, independently of the supply. (Some versions of our model allow agents to hold both assets and visit both OTC markets; see Geromichalos et al., 2023a.) As discussed, in our model an asset in larger supply implies a thicker OTC market for that asset and a higher liquidity premium, both due to an increased matching probability and a greater per-trade surplus in the OTC market. The last channel arises because, in general equilibrium, agents who opt for the less liquid, thus cheaper asset also choose to carry more money for precautionary motives.

Berentsen, Huber, and Marchesiani (2014), Mattesini and Nosal (2016), Geromichalos and Herrenbrueck (2016), Herrenbrueck (2019), and Altermatt, Iwasaki, and Wright (2023).

The rest of the paper is organized as follows. Section 2 describes the model, and Section 2.1 contains a discussion of some key modeling choices. Section 3 describes and characterizes equilibrium, and Section 4 illustrates how our model can rationalize the illiquidity of TIPS qualitatively and quantitatively. Section 5 contains a counterfactual exercise which aims to put a number on the cost of issuing TIPS for the U.S. Treasury. Section 6 extends the baseline model by adding buy-and-hold investors, and Section 7 concludes. The Appendix contains further details of the various model extensions.

2 The Model

Our model is an adaptation of Geromichalos et al. (2023a), which in turn is a hybrid of Lagos and Wright (2005) (LW henceforth) and Duffie et al. (2005). There are two important differences. First, in Geromichalos et al. (2023a) the two assets have different probabilities of default (since the research question is whether a safer asset can be endogenously more liquid); here, both assets are default free. The second difference is that the asset representing TIPS is a nominal bond that pays a face value that depends on realized inflation.[4] We now move on to the formal description of the model.

Time is discrete with an infinite horizon. Each period consists of three sub-periods. During the first sub-period, two distinct OTC asset markets open, denoted by OTC_j, $j \in \{A, B\}$. Agents who hold assets of type j can sell them for money, but they can do so only in the designated market, i.e., OTC_j. For our application, asset A will stand for (regular) Treasuries and asset B will stand for TIPS. In the second sub-period, agents visit a decentralized goods market, as is standard in LW and the subsequent literature. Trade is bilateral and the usual frictions, such as anonymity and lack of commitment, make a medium of exchange necessary. In our model only money can play that role. We refer to this market as the DM. During the third sub-period, economic activity takes place in a centralized market, which, again, is similar in spirit to the settlement market of LW. We refer to this market as the CM. There are two types of agents, consumers and producers, named by their role in the DM, and the measure of both types is normalized to the unit. All agents are infinitely lived.

[4] In related work, Geromichalos, Herrenbrueck, and Lee (2023b) also build on the framework developed in Geromichalos et al. (2023a) to examine the superior liquidity of Treasuries over corporate and municipal bonds. An important difference is that, in that paper, Treasuries have an exogenous advantage over other bond types, as they are assumed to trade in an OTC market with higher matching efficiency. That paper also studies a Cournot game between two agencies who profit from the issuance of liquid assets.

All agents discount the future between periods at rate $\beta \in (0,1)$. Consumers consume in the DM and CM and supply labor in the CM. Their preferences within a period are given by $\mathcal{U}(X,H,q) = X - H + u(q)$, where X, H stand for consumption and labor in the CM, and q stands for consumption in the DM. Producers consume in the CM and produce in both the CM and the DM. Their preferences within a period are given by $\mathcal{V}(X,H,h) = X - H - q$, where X, H are as above, and q captures units of production in the DM. We assume that u is twice continuously differentiable with $u' > 0$, $u'(0) = \infty$, $u'(\infty) = 0$, and $u'' < 0$. We let $q^* \equiv \{q : u'(q^*) = 1\}$, i.e., it denotes the optimal level of production in a bilateral DM meeting.

We now provide a more detailed description of the various sub-periods. In the third sub-period, all agents consume and produce a general good or fruit, which is also the *numeraire* good. All agents have access to a technology that transforms one unit of labor into one unit of the numeraire. Consumers (the agents who may need a medium of exchange in tomorrow's DM) choose their money holdings, which they purchase at the ongoing price φ_t (in real terms). The supply of money is controlled by a monetary authority and follows the rule $M_{t+1} = (1+\mu)M_t$, with $\mu > \beta - 1$. New money is introduced, or withdrawn if $\mu < 0$, via lump-sum transfers to consumers in the CM. Consumers can also purchase any amount of asset j at price p_j, $j \in \{A, B\}$ (in nominal terms). Each unit of asset A (Treasuries) purchased in period t's CM pays 1 dollar in the CM of period $t+1$. Each unit of asset B (TIPS) purchased in period t's CM pays $1+\pi$ dollars in the CM of period $t+1$, where π denotes the realized inflation rate between periods t and $t+1$.[5]

After making their portfolio decisions in the CM, consumers receive an idiosyncratic shock: a measure $\ell < 1$ of consumers will be active in the forthcoming DM, and we refer to them as the C-types. The remaining $1-\ell$ consumers will be referred to as the N-types ("not consuming" in the current period). Since consumers were identical when they made their portfolio choices (i.e., before the shock was realized), N-types will typically hold some money that they will not use in the current period, and C-types may be short of cash, since carrying money is costly. The OTC round of trade, which takes place before the DM, allows the reallocation of money into the hands of the types who value it most, i.e., the C-types. At the same time, N-types are happy to purchase bonds "at a good price", since they will be buying these bonds from agents who are eager to sell due to their urgent consumption need.

As we have mentioned, OTC markets are distinct: an agent who wants to sell or purchase assets is free to enter either OTC$_A$ or OTC$_B$, but she must choose one market at a time. This assumption (which we discuss in Section 2.1) has an important implication.

[5] Naturally, in a steady-state equilibrium, π equals μ and is perfectly predictable.

In equilibrium, assets will typically carry a liquidity premium, and agents are willing to pay that premium only if they can exploit the asset's liquidity services, which here simply means that they can sell the asset in its respective OTC market. Since agents can only enter one OTC market per period and anticipate having to choose between OTC$_A$ and OTC$_B$ eventually, they choose *ex-ante* (i.e., in the CM) to "specialize" in asset A or B (although, in principle, they are free to hold both assets).

Suppose that C-types and N-types have decided which market they wish to enter. (Of course, in our model these choices will be endogenous, and we discuss them in detail below.) Then, a matching function, $f(C_j, N_j)$, brings together sellers (C-types) and buyers (N-types) of assets in OTC$_j$, in bilateral fashion. We employ the specific functional form:

$$f(C_j, N_j) = \delta \left(\frac{C_j N_j}{C_j + N_j} \right)^{1-\rho} (C_j N_j)^{\rho}, \tag{1}$$

with $\delta \geq 0$, $\rho \in [0, 1]$, and $f(C_j, N_j) \leq \min\{C_j, N_j\}$. The term δ captures the efficiency of matching. A benefit of adopting this functional form is that it nests CRS as a special case. Specifically, $\rho = 0$ implies CRS in matching, while $\rho > 0$ implies IRS. Within any match in either of the OTC markets, the C-type and N-type split the surplus based on proportional bargaining (Kalai, 1977), with $\theta \in (0, 1)$ capturing the C-type's bargaining power. It is important to highlight that the matching efficiency (δ), the intensity of returns to scale (ρ), and the seller's bargaining power (θ) are *all* identical between markets A and B. In other words, our theory does not need to assume any differences in secondary market microstructure in order to capture the superior liquidity of Treasuries.

The second sub-period is the standard DM of LW. C-type consumers meet bilaterally with producers and negotiate over the terms of trade. Exchange necessitates a medium of exchange, and, as already mentioned, only money can serve that role. Since all the interesting results of the paper follow from agents' interaction in the OTC round of trade, we keep the DM as simple as possible: all C-type consumers match with a producer, and they (consumers) make a take-it-or-leave-it (TIOLI) offer.

Figure 1 summarizes the timing of the main actions in the model.

2.1 Discussion of Key Assumptions

The goal of this paper is to explain why two asset classes that are similar in every other dimension trade with different degrees of liquidity. Since in our model (and the majority of the finance literature), "liquidity" is the ease with which an asset can be traded, if needed, a necessary condition for capturing these liquidity differentials is to assume that

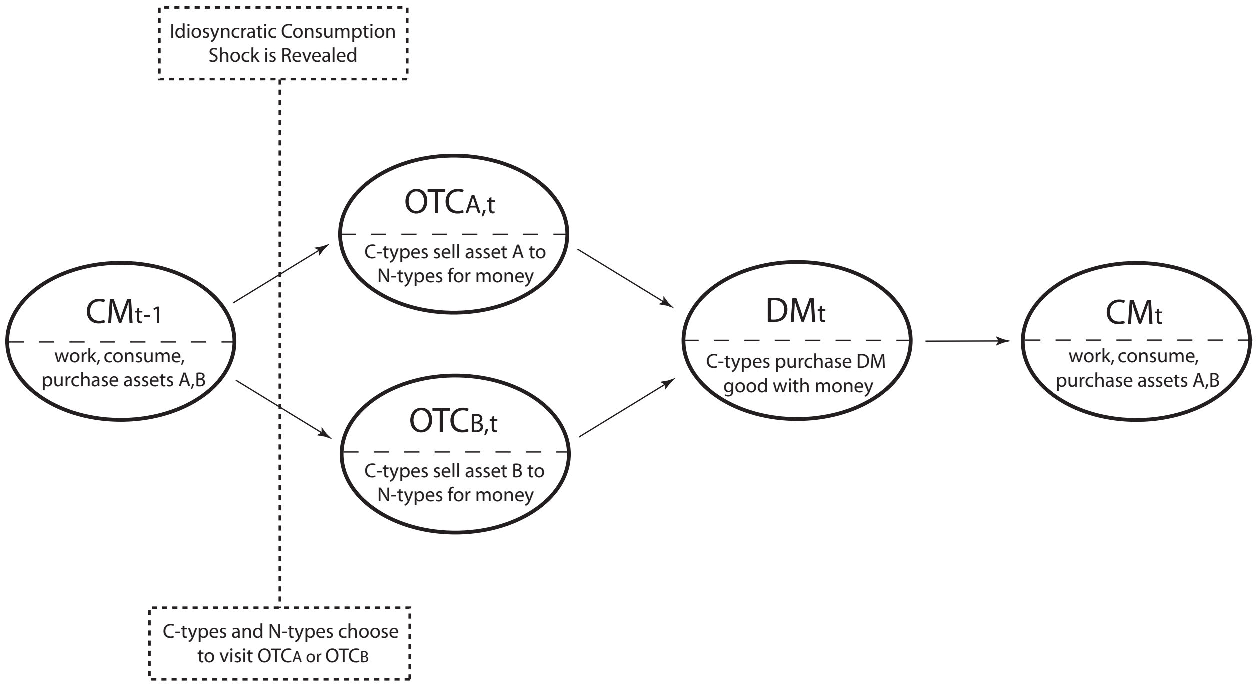

Figure 1: Timing of events.

Treasuries and TIPS are traded in *distinct* secondary markets, dubbed OTC$_A$ and OTC$_B$, respectively. (Then, the task is to study the economic forces that make one of these markets endogenously more liquid.) This assumption is also empirically relevant, since the network of dealers that specialize in trading TIPS is different from the one that specializes in Treasuries. For example, Fleckenstein et al. (2014) interviewed Treasury and TIPS traders and report: "In particular, one trader told us that there are roughly 15 dealers who were competitive in providing quotes and would be able to quickly execute purchases and sales of Treasury bonds. In contrast, the same trader indicated that there were only about five dealers who would be able to provide the same level of liquidity for TIPS." We also had our own conversation with a TIPS broker at PIMCO, who confirmed that while almost every major investment bank in the world will trade nominal Treasuries, they will not do the same with TIPS.

Generally, the literature abounds with anecdotal evidence suggesting that there is market segmentation between the secondary markets for Treasury and TIPS. Campbell et al. (2009) provide a detailed discussion about the special features of the secondary market for TIPS, including the assessment of various experts on the field. According to this discussion, Matthew Shapiro agreed that there was "substantial segmentation between the market for indexed and that for non-indexed Treasury securities." Similarly, "Michael Woodford commented on whether recent TIPS behavior indicated market segmentation.

He felt this to be the most obvious explanation." The authors conclude that yield differentials between TIPS and Treasuries are (among other reasons) "caused by liquidity premiums or technical factors that segment the bond markets."

Note that some of the related literature interprets secondary market segmentation not only as that Treasuries and TIPS trade in different secondary markets, but also as that the agents who choose to hold the two assets are inherently different. For example, Andreasen et al. (2021) assume that TIPS are more likely to enter the portfolios of "buy-and-hold" investors. As explained, our paper does not need to assume exogenous differences between the markets where the two assets trade or the investors who hold them; our mechanism can explain the TIPS illiquidity, qualitatively and quantitatively, operating only on the well-established greater supply of Treasuries. That said, we do explore a version of our model where a fraction of agents "buy and hold" TIPS, and find that the addition of this ingredient further strengthens our amplification mechanism (see Section 6). An enlightening lesson from this exercise is that the presence of buy-and-hold agents is *not sufficient* for explaining the TIPS illiquidity: if a fraction of agents buy and hold TIPS, but TIPS trade alongside Treasuries in a consolidated secondary market, the liquidity premium of TIPS will still be equal to that of nominal Treasuries (and the addition of buy-and-hold investors buys us nothing). We view this as additional justification for a claim made earlier in this section: assuming that TIPS and Treasuries trade in distinct secondary markets is a *necessary* condition for capturing the illiquidity of TIPS.

The discussion so far justifies our choice to assume that Treasuries and TIPS trade in distinct secondary markets. What is perhaps a stronger assumption is that agents can visit only one OTC market per period.[6] This assumption is not meant to be taken literally, and only intends to capture the idea that trading a particular type of asset is costly, and agents will more frequently visit the secondary market where they expect to find the best terms. In this paper, we will maintain this assumption for tractability. For the interested reader, Geromichalos et al. (2023a) also provide microfoundations (see Appendix C of that paper): specifically, the authors show that agents specializing in only one asset arises as an endogenous equilibrium outcome in a more general model where agents are given the option to trade both assets in the OTC, but choose to not exercise it.

[6] However, we do not impose any restrictions on the number of assets an agent can hold; that is, agents are free to hold both assets A and B, as well as money. The restriction is only that agents cannot visit more than one OTC markets per period. Given this restriction, and the fact that assets are typically costly to hold due to liquidity premia, agents optimally *choose* to specialize in a particular asset.

3 Analysis of the Model

3.1 Matching Probabilities

Let $e_C \in [0,1]$ and $e_N \in [0,1]$ denote the fractions of C-types and N-types, respectively, who choose to enter OTC$_A$. The measures of asset sellers and buyers in OTC$_A$ are then given by $C_A = e_C \ell$ and $N_A = e_N(1-\ell)$, and in OTC$_B$ they are $C_B = (1-e_C)\ell$ and $N_B = (1-e_N)(1-\ell)$. Given these, the matching probabilities for agents of type $i \in \{C, N\}$ in OTC$_j$, $j \in \{A, B\}$, denoted by α_{ij}, are as follows:

$$\alpha_{Cj} = \frac{f(C_j, N_j)}{C_j}, \quad \alpha_{Nj} = \frac{f(C_j, N_j)}{N_j}, \quad j = A, B. \tag{2}$$

Using the functional form (1), they can be expressed as

$$\alpha_{Cj} = \delta \frac{1}{1 + C_j/N_j}(C_j + N_j)^{\rho}, \quad \alpha_{Nj} = \delta \frac{1}{1 + N_j/C_j}(C_j + N_j)^{\rho}.$$

The matching probabilities consist of two components: (i) a market tightness term, where agents benefit from participants of the opposite type entering the market but face congestion from others of the same type, and (ii) a market size term, where agents benefit from the presence of any additional participants. The latter is what drives market concentration and is governed by the elasticity of the matching function, $\rho \in [0,1]$—absent when $\rho = 0$ (CRS), present when $\rho > 0$ (IRS), and increasing in strength as ρ rises.

3.2 Value Functions

We begin by describing the value functions in the CM. Consider a consumer entering the market with m units of money and d_j units of asset $j \in \{A, B\}$. The value function for the consumer is given by

$$\begin{aligned} W(m, d_A, d_B) = \max_{\substack{X,H,\\ \hat{m},\hat{d}_A,\hat{d}_B}} & \left\{ X - H + \beta\, \mathbb{E}_i\Big[\max\Big\{ \Omega_{iA}\Big(\hat{m}, \hat{d}_A, \hat{d}_B\Big), \Omega_{iB}\Big(\hat{m}, \hat{d}_A, \hat{d}_B\Big) \Big\}\Big] \right\} \\ \text{s.t. } & X + \varphi(\hat{m} + p_A \hat{d}_A + p_B \hat{d}_B) = H + \varphi(m + d_A + (1+\pi)d_B + \mu M), \end{aligned}$$

where variables with hats indicate next-period choices, $\mathbb{E}_i$ is the expectation operator with respect to the idiosyncratic consumption shock, and Ω_{ij} denotes the value function for a consumer of type $i \in \{C, N\}$ entering OTC$_j$, $j \in \{A, B\}$. The price of money is in terms of the general good, while the price of bonds is in nominal terms. At the optimum, X

and H are indeterminate, but their difference is not. Substituting $X - H$ from the budget constraint into W yields

$$
\begin{aligned}
W(m, d_A, d_B) = \varphi(m + d_A + (1+\pi)d_B + \mu M) \\
+ \max_{\hat{m}, \hat{d}_A, \hat{d}_B} \Bigg\{ & - \varphi(\hat{m} + p_A \hat{d}_A + p_B \hat{d}_B) \\
& + \beta \ell \cdot \max \Big\{ \Omega_{CA}\Big(\hat{m}, \hat{d}_A, \hat{d}_B\Big), \Omega_{CB}\Big(\hat{m}, \hat{d}_A, \hat{d}_B\Big) \Big\} \\
& + \beta(1-\ell) \cdot \max \Big\{ \Omega_{NA}\Big(\hat{m}, \hat{d}_A, \hat{d}_B\Big), \Omega_{NB}\Big(\hat{m}, \hat{d}_A, \hat{d}_B\Big) \Big\} \Bigg\},
\end{aligned} \tag{3}
$$

where we use the fact that the representative buyer becomes a C-type with probability ℓ to replace the expectation operator. As is standard in models that build on LW, the agent's optimal choice does not depend on the current state due to the quasi-linear preferences, and the CM value function is linear.

Naturally, producers will not want to leave the CM with positive money and bond holdings. Thus, when entering the CM, they will only hold money received as payment in the preceding DM, and their CM value function is given by

$$W^P(m) = \max_{X,H} \left\{ X - H + V^P \right\} \quad \text{s.t.} \;\; X = H + \varphi m,$$

where V^P denotes the seller's value function in the forthcoming DM. Substituting $X - H$ from the budget constraint shows that W^P is linear: $W^P(m) = \varphi m + V^P$.

Next, we describe the OTC value functions. Let ξ_j denote the amount of money transferred to the C-type and χ_j the amount of asset j transferred to the N-type in a typical match in OTC$_j$, $j \in \{A, B\}$. Using the matching probabilities α_{ij} defined in (2), the value function for an agent of type i entering OTC$_j$, Ω_{ij}, is given by

$$
\begin{aligned}
\Omega_{CA}(m, d_A, d_B) &= \alpha_{CA} V\left(m + \xi_A, d_A - \chi_A, d_B\right) + \left(1 - \alpha_{CA}\right) V(m, d_A, d_B), \\
\Omega_{CB}(m, d_A, d_B) &= \alpha_{CB} V\left(m + \xi_B, d_A, d_B - \chi_B\right) + \left(1 - \alpha_{CB}\right) V(m, d_A, d_B), \\
\Omega_{NA}(m, d_A, d_B) &= \alpha_{NA} W\left(m - \xi_A, d_A + \chi_A, d_B\right) + \left(1 - \alpha_{NA}\right) W(m, d_A, d_B), \\
\Omega_{NB}(m, d_A, d_B) &= \alpha_{NB} W\left(m - \xi_B, d_A, d_B + \chi_B\right) + \left(1 - \alpha_{NB}\right) W(m, d_A, d_B),
\end{aligned}
$$

where V denotes the consumer's value function in the DM. Note that N-type consumers proceed directly to the next period's CM.

Lastly, consider the value functions in the DM. Let q denote the quantity of special goods traded and τ the total payment in money. The value function for a consumer entering this market is given by

$$V(m, d_A, d_B) = u(q) + W(m - \tau, d_A, d_B),$$

and the DM value function for a seller entering with no money or assets is given by

$$V^P = -q + W^P(\tau).$$

3.3 Terms of Trade

First, consider a DM meeting between a C-type consumer with portfolio (m, d_A, d_B) and a producer. The two parties bargain over a quantity q to be produced by the producer and a cash payment τ from the consumer. The consumer makes a TIOLI offer to maximize her surplus, subject to the producer's participation constraint and her own cash constraint:

$$\max_{\tau,q} \{u(q) + W(m - \tau, d_A, d_B) - W(m, d_A, d_B)\}$$
$$\text{s.t. } -q + W^P(\tau) - W^P(0) = 0 \text{ and } \tau \leq m.$$

The linearity of the value functions W and W^P reduces the C-type consumer's surplus to $u(q) - \varphi\tau$ and the producer's surplus to $-q + \varphi\tau$. This implies that the bargaining solution must satisfy $q = \varphi\tau$, meaning the producer requires q/φ units of money to produce q of special goods. If the agent has enough money to purchase the optimal quantity, i.e., $m \geq q^*/\varphi$, then q^* is produced; otherwise, φm is produced. Let $m^* \equiv q^*/\varphi$ be the amount of money needed for the first-best quantity, q^*. The solution is then expressed as

$$\tau(m) = \min\{m^*, m\}, \tag{4}$$
$$q(m) = \min\{q^*, \varphi m\}. \tag{5}$$

Note that, since $\mu > \beta - 1$, the cost of carrying money is positive, so consumers never leave the CM or exit the OTC round of trade with an amount exceeding m^*. Thus, the relevant case is when the consumer's cash constraint binds: $\tau(m) = m$ and $q(m) = \varphi m$.

Now, we describe the terms of OTC trade. We focus on OTC$_B$; for OTC$_A$, everything is the same except that 1 replaces $1 + \pi$. Consider a meeting in OTC$_B$ between a C-type consumer with portfolio (m, d_A, d_B) and an N-type consumer with portfolio $(\widetilde{m}, \widetilde{d}_A, \widetilde{d}_B)$. The two parties negotiate a transfer of money, ξ_B, to the C-type and an amount asset B, χ_B, to the N-type. The surplus is split based on proportional bargaining, with $\theta \in (0, 1)$ denoting the C-type's bargaining power, subject to the C-type's asset holding constraint and the N-type's money holding constraint. The surpluses for both agents are given by

$$S_{CB} = V(m + \xi_B, d_A, d_B - \chi_B) - V(m, d_A, d_B),$$
$$S_{NB} = W(\widetilde{m} - \xi_B, \widetilde{d}_A, \widetilde{d}_B + \chi_B) - W(\widetilde{m}, \widetilde{d}_A, \widetilde{d}_B),$$

where the linearity of the value function W reduces S_{CB} to $u(q(m+\xi_B)) - u(q(m)) - \varphi(1+\pi)\chi_B$ and S_{NB} to $-\varphi\xi_B + \varphi(1+\pi)\chi_B$. The bargaining problem is then described by

$$\max_{\xi_B, \chi_B} S_{CB} \quad \text{s.t.} \quad S_{CB} = \frac{\theta}{1-\theta} S_{NB}, \quad \chi_B \leq d_B, \quad \text{and} \quad \xi_B \leq \widetilde{m}.$$

The Kalai constraint implies

$$\varphi(1+\pi)\chi_B = (1-\theta)(u(q(m+\xi_B)) - u(q(m)) + \theta\varphi\xi_B \equiv \sigma(\xi_B),$$

meaning that the asset seller has to give up $\sigma(\xi_B)/(\varphi(1+\pi))$ units of asset B for ξ_B units of money, with more assets required for larger amounts of money, as indicated by $\sigma' > 0$. Since carrying money is costly, a consumer will never bring more than m^* from the CM and will want to acquire the amount needed to reach m^*, namely, $m^* - m$. If she has enough assets to acquire $m^* - m$, i.e., $d_B \geq \sigma(m^* - m)/(\varphi(1+\pi))$, then she acquires $m - m^*$; otherwise, she sells all her assets and acquire $\sigma^{-1}(\varphi(1+\pi)d_B)$ units of money.[7] The bargaining solution is thus given by

$$\xi_B(m, d_B) = \min\left\{m^* - m, \sigma^{-1}(\varphi(1+\pi)d_B)\right\}, \tag{6}$$
$$\chi_B(m, d_B) = \min\left\{\frac{\sigma(m^* - m)}{\varphi(1+\pi)}, d_B\right\}. \tag{7}$$

Note that the solution does not depend on the N-type's portfolio, as we assume that the N-type's money holding constraint does not bind.

3.4 Optimal Portfolio Choice

Agents choose their optimal portfolio in the CM independently of their trading histories in previous markets, as is standard in models that build on LW. Additionally, they decide

[7] Note that this discussion assumes the N-type's money holdings do not constrain the trade, effectively ignoring the constraint $\xi_B \leq \widetilde{m}$ in the bargaining problem. A sufficient condition for this assumption is that the combined money holdings of the C-type and N-type allow the C-type to purchase the first best quantity, q^*, i.e., $m + \widetilde{m} \geq m^*$. This condition is satisfied in equilibrium, provided that inflation is not too high, ensuring all agents carry at least half of the first-best amount of money. We confirm that this condition holds under our calibrated parameters.

which OTC market to enter once their type is revealed. The objective function is derived by substituting the value functions into (3) and retaining only the terms that depend on the choice variables:

$$
\begin{aligned}
J(\hat{m}, \hat{d}_A, \hat{d}_B) = &- \varphi(\hat{m} + p_A \hat{d}_A + p_B \hat{d}_B) + \beta\hat{\varphi}(\hat{m} + \hat{d}_A + (1+\pi)\hat{d}_B) \\
&+ \beta\ell(u\left(\hat{\varphi}\hat{m}\right) - \hat{\varphi}\hat{m} + \max\{\alpha_{CA}S_{CA}, \alpha_{CB}S_{CB}\}),
\end{aligned} \tag{8}
$$

The first two terms represent the cost of choosing a portfolio and its payout in the next period's CM. The portfolio also offers liquidity benefits if an agent becomes a C-type: they can use their money from the CM to purchase special goods in the DM and gain additional benefits by selling assets for more cash in the OTC markets. The $\max$ operator indicates the decision of which market to enter. The objective function excludes terms related to being an N-type consumer, as they do not consume in the DM and as the terms of OTC trade are independent of the N-type's portfolio. Recall that since agents can enter only one OTC market per period, and assets typically trade at a liquidity premium, agents will "specialize" ex ante in either asset A or B, holding only one of these asset types. Importantly, an agent's specialization decision can (and will) affect their demand for money. For example, an agent who chooses to hold an asset that she deems less liquid, i.e., harder to sell in the OTC, will typically choose to carry more money (as insurance against the higher liquidity risk).

3.5 Equilibrium

In the steady-state equilibrium conditions, we summarize the cost of holding money via $i \equiv (1+\mu)/\beta - 1$, which can also be interpreted as the nominal yield on a completely illiquid asset. (Thus, i should not be thought of as representing, for instance, the yield on T-bills; see Geromichalos and Herrenbrueck, 2022 and Herrenbrueck and Wang, 2023.)

A steady-state equilibrium consists of the equilibrium prices $\{\varphi, p_A, p_B\}$, quantities $\{q_{0A}, q_{0B}, q_{1A}, q_{1B}\}$, and entry choices $\{e_C, e_N\}$. Here, q_{kj} represents the amount of special goods purchased by a C-type agent in the DM, where $k \in \{0, 1\}$ indicates whether the agent traded ($k = 1$) or did not trade ($k = 0$) in the preceding OTC market, and $j \in \{A, B\}$ indicates the asset the agent specializes in. Given the consumer's TIOLI offers in the DM, as shown in (4) and (5), q_{0j} equals the real balances held by agents specializing in asset j.

Once the quantities and entry choices are determined, the equilibrium prices immediately follow. The price of money comes from the money market clearing condition:

$$\varphi M = e_C q_{0A} + (1 - e_C) q_{0B}, \tag{9}$$

The asset prices are derived from the optimal asset portfolio choice:

$$p_A = \frac{1}{1+i}(1 + L_A), \quad p_B = \frac{1+\pi}{1+i}(1 + L_B), \tag{10}$$

where the liquidity premium of asset j, denoted as L_j, is equal to

$$L_j = \ell \alpha_{Cj} \frac{\theta}{w(q_{1j})}(u'(q_{1j}) - 1), \quad j = A, B, \tag{11}$$

with $w(q_{1j}) \equiv 1 + (1-\theta)(u'(q_{1j}) - 1) \in [1, \infty)$. An asset is valued for its payout at maturity and the additional benefit arising from the possibility of selling it in the OTC market. The fundamental value of an asset is defined as the equilibrium price that would emerge if this possibility of OTC trading were eliminated, which is $1/(1+i)$ for asset A and $(1+\pi)/(1+i)$ for asset B. The liquidity premium of asset j, L_j, is defined as the percentage difference between its price and fundamental value and is the product of four terms: (i) ℓ, the probability that an agent turns out to be a C-type (an asset seller), (ii) α_{Cj}, the probability of matching in OTC$_j$ provided that the agent is a C-type, (iii) $\theta/w(q_{1j})$, the effective bargaining share for a C-type in OTC trade, and (iv) $u'(q_{1j}) - 1$, the marginal DM trade surplus from the last unit of real balances.[8] The liquidity premium will be zero either when no buyers enter OTC$_j$ (i.e., $\alpha_{Cj} = 0$) or when assets are so plentiful that $q_{1j} = q^*$. In the former case, assets are illiquid; in the latter case, assets are still liquid, but their liquidity is inframarginal and does not affect the price.

The quantities and entry choices satisfy the following set of equilibrium conditions. First, the money demand equation arises from the optimal money holding decision:

$$i = \ell\left(1 - \alpha_{Cj}\frac{\theta}{w(q_{1j})}\right)(u'(q_{0j}) - 1) + \ell\alpha_{Cj}\frac{\theta}{w(q_{1j})}(u'(q_{1j}) - 1), \quad j = A, B. \tag{12}$$

Notice that, as discussed earlier, agents who specialize in different assets typically carry different amounts of money. Second, the OTC trading protocol $q_{1j} = q_{0j} + \varphi\xi_j$, combined with the bargaining solution (as in (6) and (7)) and the market clearing conditions ($A =$

[8] The effective bargaining share is an endogenous object, whose value lies within the range of $(0, \theta]$. It consists of two terms: θ, the bargaining power of C-types in OTC trade, and $1/w(q_{1j})$, which equals $\partial\xi_j/\partial d_j$ divided by the face value of asset j, representing the marginal amount of money obtained from selling the last unit of asset j holdings per unit of face value. Note that $1/w(q_{1j}) < 1$ unless $q_{1j} = q^*$, indicating that, due to $\theta < 1$, agents receive real balances worth less than the real value of the assets they sell.

$e_C d_A$, $B = (1 - e_C)d_B$, and $\varphi M = e_C q_{0A} + (1 - e_C)q_{0B}$), leads to

$$q_{1A} = \min\left\{q^*,\, q_{0A} + \frac{1}{\theta}\frac{A}{M}\frac{e_C q_{0A} + (1 - e_C)q_{0B}}{e_C} - \frac{1-\theta}{\theta}(u(q_{1A}) - u(q_{0A}))\right\}, \tag{13}$$

$$q_{1B} = \min\left\{q^*,\, q_{0B} + \frac{1}{\theta}\frac{(1+\pi)B}{M}\frac{e_C q_{0B} + (1 - e_C)q_{0B}}{1 - e_C} - \frac{1-\theta}{\theta}(u(q_{1B}) - u(q_{0B}))\right\}. \tag{14}$$

The last two equilibrium conditions concern the agents' optimal entry decisions in the OTC market. As discussed following (8), since the OTC markets are segmented and assets typically trade at a premium, agents specialize ex ante in either asset A or B. This implies that an agent's portfolio choice—especially which asset to hold—is intertwined with the decision of which OTC market to enter if the agent turns out to be a C-type. Therefore, when making entry decisions, C-types must weigh not only the expected surplus of entering a market but also the cost of holding money and the associated asset. The optimal entry of C-type consumers is given by

$$e_C = \begin{cases} 1, & \widetilde{S}_{CA} > \widetilde{S}_{CB} \\ 0, & \widetilde{S}_{CA} < \widetilde{S}_{CB} \\ \in [0,1], & \widetilde{S}_{CA} = \widetilde{S}_{CB}\,, \end{cases} \tag{15}$$

where

$$\begin{aligned} &\widetilde{S}_{Cj} = -iq_{0j} - L_j((1-\theta)(u(q_{1j}) - u(q_{0j})) + \theta(q_{1j} - q_{0j})) + \ell(u(q_{0j}) - q_{0j} + \alpha_{Cj}S_{Cj}), \\ &S_{Cj} = \theta(u(q_{1j}) - u(q_{0j}) - q_{1j} + q_{0j}). \end{aligned}$$

In contrast, N-types make entry decisions ex post, based solely on the expected surplus from entering a market, since, unlike C-types, their portfolios do not influence the terms of OTC trade (recall footnote 7). For N-type consumers, optimal entry is given by

$$e_N = \begin{cases} 1, & \alpha_{NA}S_{NA} > \alpha_{NB}S_{NB} \\ 0, & \alpha_{NA}S_{NA} < \alpha_{NB}S_{NB} \\ \in [0,1], & \alpha_{NA}S_{NA} = \alpha_{NB}S_{NB}\,, \end{cases} \tag{16}$$

where

$$S_{Nj} = (1-\theta)(u(q_{1j}) - u(q_{0j}) - q_{1j} + q_{0j}).$$

We now define a steady-state equilibrium of the model.

Definition 1. For given asset supplies $(A, B) \in \mathbb{R}^2_+$, a steady-state equilibrium is a list

$\{\varphi, p_A, p_B, q_{0A}, q_{1A}, q_{0B}, q_{1B}, e_C, e_N\}$ that solves (9), (10), (12), (13), (14), (15), and (16), with the matching probabilities given by (2).

We construct equilibria as fixed points of e_N. Specifically, for a given e_N, we compute agents' optimal portfolio choices and define $G(e_N) \equiv (\alpha_{NA}S_{NA} - \alpha_{NB}S_{NB})/(\alpha_{NA}S_{NA} + \alpha_{NB}S_{NB})$, which measures the relative benefit for an individual N-type in choosing OTC$_A$ over OTC$_B$, taking as given the decisions of other agents. Starting with $e_N = A/(A + (1+\pi)B)$—the equilibrium entry in a symmetric equilibrium where $e_C = e_N$, $q_{0A} = q_{0B}$, and $q_{1A} = q_{1B}$ under CRS OTC matching—we adjust e_N according to the sign of $G(e_N)$: increasing (decreasing) it if $G(e_N) > 0$ (< 0), until we reach an equilibrium (robust to small trembles) where $G(e_N) = 0$. If multiple robust equilibria exist, we find the one closest to the starting point (in the direction of the sign of $G(e_N)$).[9]

4 Quantitative Analysis: Rationalizing TIPS Illiquidity

We calibrate the model to the Treasury markets for Treasuries and TIPS, using data on the supply of both assets. We then test whether the calibrated model can quantitatively match the observed yields of these assets, as well as their spread. Using the utility function $u(q) = q^{1-\sigma}/(1-\sigma)$, we calibrate ten parameters listed in Table 1. Some parameters have direct empirical counterparts, while others do not. For the latter, we identify the combinations that precisely match the model to the observed yields and spread.

For the monetary aggregate, we use the MZM money stock from the Federal Reserve Bank of St. Louis. As shown in (13) and (14), only the assets' relative supplies to the money supply matter, so we normalize M to 1. Data on Treasuries and TIPS supplies come from the U.S. Treasury Monthly Statement of the Public Debt. From July 2009 to February 2020 (post-Great Recession to pre-Covid, according to the business cycle dating by the National Bureau of Economic Research), the average outstanding amounts of Treasuries and TIPS of 10-year maturity, relative to the monetary aggregates, are 0.153 and 0.0425, respectively.[10] Kim, Walsh, and Wei (2019), in the FEDS Notes of the Federal Reserve Board of Governors, estimate the average expected inflation as 2.2996% during the same period, which equals the money growth rate and the realized inflation in the steady-state equilibrium of the model; thus we set μ to 2.2996%. The discount rate is set to 1/1.03, consistent with a 3% annual real return from Bethune, Choi, and Wright (2020).

[9] Specifically, we adjust e_N in increments (decrements) of 10^{-5} when $G(e_N) > 0$ (< 0). Once the sign of $G(e_N)$ reverses, we apply the bisection method to find the value of e_N such that $G(e_N) = 0$. If a corner ($e_N = 0$ or 1) is reached without a change in the sign of $G(e_N)$, we identify that corner as the robust equilibrium.

[10] We focus on 10-year maturities to be consistent with the yield decomposition data we use.

	Description	Value
A	supply of Treasuries	0.1530
B	supply of TIPS	0.0425
M	money supply	1
β	discount rate	1/1.03
μ	growth rate of money supply	2.2996%
ℓ	fraction of C-type agents	0.5
σ	elasticity of marginal utility	0.34
θ	relative bargaining power of C-types	See Table 3
δ	matching efficiency in the OTC markets	See Table 3
ρ	elasticity of the OTC matching function	See Table 3

Table 1: Key parameter values.

We set $\ell = 0.5$ for symmetry (equal numbers of potential buyers and sellers in the OTC markets) and $\sigma = 0.34$ to match the slope of the empirical U.S. money demand, using the procedure of Rocheteau, Wright, and Zhang (2018).

For the remaining three parameters (θ, δ, ρ), we identify combinations that match the yields of Treasuries and TIPS, as well as the spread between them, to their empirical counterparts in Table 2. The data, also sourced from Kim et al. (2019), pertain to 10-year Treasuries and TIPS, with Treasury yields adjusted for the inflation risk premium, as our model abstracts from inflation uncertainty. In the model, yields for Treasuries and TIPS are given by $j_A = i - L_A$ and $j_B = i - \pi - L_B$, respectively, with the TIPS (il)liquidity premium relative to Treasuries given by $L_A - L_B$. With an extra degree of freedom, we vary θ and, for each value of θ, identify a (δ, ρ) combination that aligns the model with the data. The results are presented in Table 3.[11] For $\theta < 0.7$, no exact match is possible, which is quite intuitive. The bargaining power of asset sellers in OTC trade symmetrically scales the liquidity premia of both assets, as shown in (11). Therefore, a sufficiently large θ is needed to match the sizable liquidity premia of Treasuries and TIPS ($L_A = 2.7324\%$ and $L_B = 2.4931\%$, as implied by Tables 1 and 2). At higher values of θ, the model can match the data, up to $\theta \to 1$ in the limit; however, θ must remain below 1 to preserve a meaningful entry decision for asset buyers.

Overall, our model requires only a minimal degree of IRS in OTC matching, with ρ

[11] For all (θ, δ, ρ) combinations, $q_{0j}/q^* > 0.5$ for $j = A, B$. For instance, with $(\theta, \delta, \rho) = (0.8, 1.6392, 0.0089)$, we have $q_{0A}/q^* = 0.6612$ and $q_{0B}/q^* = 0.6678$. Given the consumers' TIOLI offers in the DM, q_{0j} equals the real balances held by agents specializing in asset j. Thus, $q_{0j}/q^* > 0.5$ implies $m + \widetilde{m} \geq m^*$, confirming our earlier claim (in footnote 7) that N-type's money holdings never bind in OTC trade.

Target	Data
TIPS (il)liquidity premium (relative to Treasuries)	0.2393%
average yields of TIPS	0.5069%
average yields of Treasuries	2.5672%

Table 2: Targeted empirical moments.

θ	δ	ρ
0.7	1.8853	0.0092
0.8	1.6392	0.0089
0.9	1.4477	0.0086

Table 3: Combinations of (δ, ρ) that exactly match the model to the observed TIPS (il)liquidity premium and yields of TIPS and Treasuries, for given values of θ.

below 0.01, virtually in the neighborhood of CRS. Nevertheless, ρ must still be positive; otherwise, $L_A = L_B$, as illustrated in Figure 2, which plots the liquidity premia of assets A and B as the supply of A varies while the supply of B remains fixed. As the supply of A increases, more agents seek to trade in OTC$_A$, anticipating a larger potential trade surplus. Under CRS matching, however, agents avoid markets crowded with agents of the same type, as congestion lowers their chances of finding a trading partner. This results in constant market tightness ($e_N/e_C = 1$), equal across both markets, leading to $q_{1A} = q_{1B}$, $\alpha_{CA} = \alpha_{CB}$, and $L_A = L_B$, as shown in the right panel of the figure.

In contrast, IRS ($\rho > 0$) promote market concentration; that is, agents prefer markets with more participants. Crucially, when agents are attracted to OTC$_A$ by the increase in the supply of A, the response of asset buyers is more sensitive than that of sellers, as buyers make entry decisions ex post. Consequently, a higher supply of A increases market tightness and the sell-probability in OTC$_A$, implying a greater ex-ante willingness to pay for asset A due to its liquidity. This liquidity advantage is reinforced by a general-equilibrium mechanism: recognizing that matching in OTC$_B$ will be more challenging, agents specializing in asset B bring more money, relative to those specializing in asset A. This reduces their reliance on an additional liquidity boost and diminishes the potential trade surplus in OTC$_B$, thereby reinforcing the dominance of OTC$_A$. Even a minimal degree of IRS amplifies these effects, leading to upward-sloping asset demand curves, as shown in the left panel of the figure, and making assets in greater supply more liquid.

In Table 3, the levels of δ and ρ required to match the data adjust as θ varies. As noted, θ scales the liquidity premia of the two assets proportionally, influencing both their size

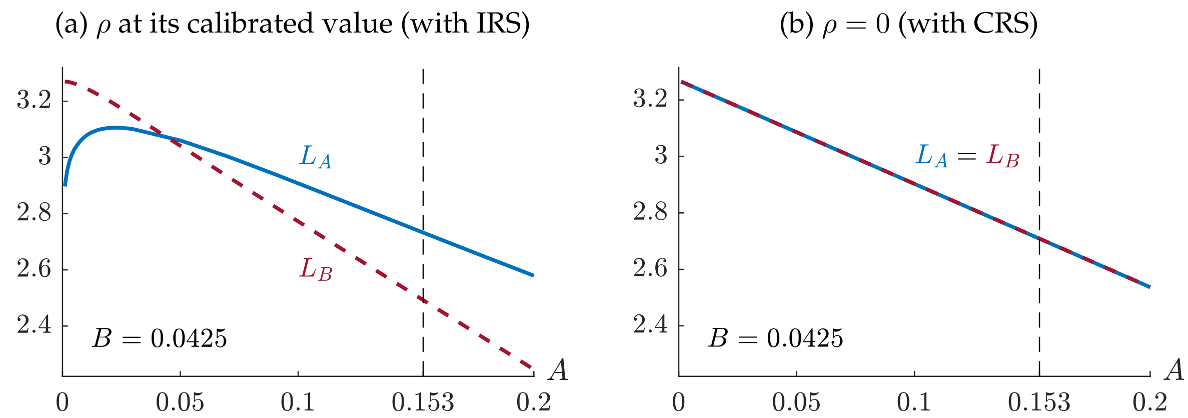

Figure 2: Liquidity premia of assets A and B, with varying supply of A and fixed supply of B. The vertical dashed line marks the supply of asset A from Table 1, while the supply of asset B is shown in the bottom left corner. In the left panel, other parameters are set to their calibrated values in Tables 1 and 3, with $\theta = 0.8$. In the right panel, ρ is set to 0, with all other parameters unchanged.

and the liquidity differential. Similarly, δ impacts the liquidity premia by affecting the matching efficiency in both OTC markets.[12] IRS ($\rho > 0$) facilitates an equilibrium where $L_A > L_B$, with a higher ρ increasing the liquidity differential. A higher θ corresponds to lower values of δ and ρ, as it reduces the burden for these parameters to fit the data. This exercise demonstrates that a reasonable parametrization of our model can quantitatively explain the observed levels and differential of Treasury and TIPS yields.

5 Counterfactual Exercise: the Cost of Issuing TIPS

When the Treasury introduced TIPS in 1997, their intent was to offer a new type of security that would be appealing to investors by eliminating inflation risk, and, hence, cheaper for the Treasury. Several reports published in the months that preceded the introduction of TIPS talk about how fantastic this new product would be, and some calculated the interest rate savings the Treasury would achieve when the new securities would hit the market.[13] With hindsight, the introduction of TIPS did not save taxpayers billions of dollars, as

[12] For $\delta > 1$, we truncate the matching function so that $f(C, N) \leq \min\{C, N\}$ and $\alpha_{ij} \leq 1$ for $i = C, N$ and $j = A, B$. For all (θ, δ, ρ) combinations, $\alpha_{ij} < 1$. For instance, with $(\theta, \delta, \rho) = (0.8, 1.6392, 0.0089)$, we have $\alpha_{CA} = 0.8273$, $\alpha_{CB} = 0.774$, $\alpha_{NA} = 0.8083$, and $\alpha_{NB} = 0.8429$.

[13] For example, the journalists of a *Wall Street Journal* article (on January 21, 1997; available at https://www.wsj.com/articles/SB853762604845980000) interviewed Deputy Secretary of the Treasury Lawrence Summers and reported that "Mr. Summers trumpeted the new product as 'the most significant innovation in federal debt-management policy in the last generation' saying that the new notes will 'reduce federal borrowings costs'." (See also the epigraph to this paper.) For an estimation of the interest rate savings the Treasury would achieve by issuing TIPS, see Campbell and Shiller (1996).

had been predicted, but in fact led to the opposite outcome because TIPS turned out to be significantly less liquid than nominal Treasuries. This spurred a debate over whether TIPS had been a mistake, and whether enough attention had been paid to their liquidity. Campbell et al. (2009) report:

> "As a long-time advocate of inflation-linked bonds, Alan Blinder had been excited when Campbell and Shiller's 1996 paper put an actual number on the likely interest rate savings to the Treasury [by issuing TIPS]. That paper, he recalled, said that TIPS should be cheaper for the Treasury because they were less risky to bondholders and would therefore pay a lower rate of return. In reality, they have not paid a lower rate, which, Blinder reasoned, was due to their lesser liquidity compared with nominal bonds. He wondered whether the main message of the paper was that economists have been focusing too much on risk and not enough on liquidity."
>
> "Gregory Mankiw addressed Alan Blinder's comment that a major argument for the creation of TIPS had been their lower cost of financing for the Treasury. He wondered whether that argument had been the primary one, and, if it had and now turned out to be wrong, whether Blinder felt that TIPS had been a mistake and should be phased out. Blinder responded that it had been the primary argument and that TIPS were a mistake from that perspective, but that TIPS should not therefore disappear, because they still provide a low-risk investment vehicle for investors, albeit at a cost to taxpayers."

And Fleckenstein et al. (2014):

> "Finally, we show that TIPS are almost always too 'cheap' relative to Treasury bonds. An immediate implication of this is that the Treasury could have reduced the cost of the public debt by issuing only nominal bonds [...] Our results suggest that the policy may be far more costly than previously recognized. This is because the Treasury [...] *leaves billions of dollars on the table* by issuing securities that are not as highly valued by the market as nominal Treasury bonds." (Emphasis added.)

The last two quotes motivate the counterfactual exercise performed in this section. Specifically, we use our model to answer two policy-relevant questions: a) how many (billions of) dollars does the Treasury leave on the table "by issuing securities that are not as highly valued by the market as nominal Treasury bonds"; and, b) should the Treasury

			Liquidity Premium	Yield		Gain
θ	δ	ρ	$L_A = L_B$	Treasuries	TIPS	(USD Millions)
0.7	1.8853	0.0092	2.6987%	2.5997%	0.2934%	694.19
0.8	1.6392	0.0089	2.6985%	2.5999%	0.2936%	688.26
0.9	1.4477	0.0086	2.6979%	2.6005%	0.2942%	669.97

Table 4: Counterfactual liquidity premia (fourth column) and yields of Treasuries and TIPS (fifth and sixth columns) in a model with a consolidated secondary asset market for combinations of (θ, δ, ρ) in Table 3. The last column shows how much the U.S. Treasury could save on the issuance of Treasuries and TIPS in 2019 U.S. dollars, provided the secondary markets are consolidated.

phase out TIPS? The answers to these two questions are intertwined. Through the lens of our model, TIPS are less liquid because they trade in segmented secondary markets, which gives a liquidity advantage to their non-inflation indexed counterparts that come at a much larger supply. The Treasury could of course save taxpayers' money by discontinuing the issuance of a type of security that markets clearly value less than Treasuries. But there is another solution: the Treasury could take measures that reduce secondary market segmentation and ensure that TIPS trade alongside nominal Treasuries in *consolidated* secondary markets.

Thus, in regard to the second question posed above, our model suggests that *it is possible* for the Treasury to continue issuing this desirable "low-risk investment vehicle" (in Alan Blinder's words) without imposing a "cost to taxpayers". As for the first question—putting a number on the Treasury's cost of issuing TIPS—we can reorient it by asking, "How much would the Treasury save if they could ensure that TIPS and nominal Treasuries trade in a consolidated secondary market?" In what follows, we detail our model's answer to this policy-relevant question. The section concludes with a short discussion of how such consolidation could be achieved in reality.

We perform the following counterfactual exercise: we ask what the gains in terms of higher liquidity premia (lower borrowing rates) would be from merging the secondary markets OTC$_A$ and OTC$_B$ into a single consolidated market. The details of this version of the model can be found in the appendix, and the results are summarized in Table 4. The first three columns of the table display combinations of (θ, δ, ρ) under which the benchmark model (with segmented secondary markets) produces the observed yields and liquidity premia of Treasuries and TIPS (see Tables 2 and 3). In our counterfactual exercise, we calculate the yields and liquidity premia in the consolidated secondary market model for each of these combinations of parameters. With a consolidated market, the two assets become perfect substitutes in the provision of liquidity; this simply means that their liquidity premia will be equal, $L_A = L_B$, and these premia are reported in the fourth column of the table. The counterfactual Treasury and TIPS yields are shown in the fifth and sixth columns, with their differential reflecting expected inflation.

Compared to the benchmark, in the counterfactual model, the Treasury yield is higher by roughly 3.3 basis points, while the TIPS yield is lower by about 21.3 basis points. Clearly, consolidating the secondary markets implies that the Treasury makes a loss (compared to the benchmark model) on the issuance of non-indexed securities, which have now lost their liquidity advantage over TIPS. However, the increase in the Treasury yield is an order of magnitude smaller than the decrease in the TIPS yield. This result is primarily driven by changes in market tightness and their effects on the sell-probability. In the benchmark model, IRS promote market concentration in markets for assets with a larger supply, leading to $0.5 < e_C < e_N$, due to the flexibility of N-types' ex-post entry decisions (as explained in detail in Section 4). Moreover, market tightness in OTC$_A$ and OTC$_B$ is given by $t_A \equiv e_N(1-\ell)/(e_C\ell)$ and $t_B \equiv (1-e_N)(1-\ell)/((1-e_C)\ell)$, respectively, whereas in the consolidated OTC market model, the (common) market tightness is fixed at $t \equiv (1-\ell)/\ell$. In all, we have $t_A > t > t_B$, but the size of the differential $|t_A - t|$ is far smaller than $|t_B - t|$. Therefore, consolidating the OTC markets causes the sell-probability, and the resulting liquidity premium, for Treasuries to decrease by a significantly smaller margin than the corresponding increase for TIPS.[14]

Consequently, the savings from consolidating the markets on the issuance of TIPS, due to their higher price, outweigh the loss on the issuance of Treasuries resulting from their lower price. The net gain is given by $(\widetilde{p}_A - p_A)A + (\widetilde{p}_B - p_B)B$, where $\widetilde{p}_A < p_A$ and $\widetilde{p}_B > p_B$, with the tilde indicating counterfactual variables. The last column of Table 4 presents the estimated gain in 2019 U.S. dollars. According to our calibrated model, the U.S. Treasury would save between 669.97 and 694.19 million dollars per year, on average, on the given issuance of Treasuries and TIPS (in their 10-year maturity segment) during the post-Great Recession to pre-Covid period, if the two assets were traded in a consolidated OTC market. Notice that while the range of parameters over which we perform our counterfactual exercise (see Table 4) is rather wide, the range of the estimated savings from consolidating the two secondary markets is quite narrow (approximately equal to 700 million dollars per year). We view this as an indication that our quantitative exercise

[14] There is also a market size effect: both assets A and B benefit from the increased number of participants following the consolidation of the markets. However, quantitatively, this benefit for asset A is smaller than the loss from lower market tightness, as OTC$_A$ already had the majority of participants due to asset A's large supply. This, in turn, means that the benefit from the market size effect is greater for asset B, as OTC$_B$ had fewer participants than OTC$_A$ before the consolidation.

is robust and that the reported numbers are informative.

While the goal of this section is to use our model to *put a number* on the cost of issuing TIPS, thus informing the recent debate, it is natural to ask what could be done in reality to facilitate the consolidation of secondary markets, which is the model's main prescription. The answer to this question is not obvious or easy. For one, it is unlikely that the Treasury has full control over the organization of the secondary market(s) where its securities trade. In Section 2.1, we reported that the network of dealers that specialize in trading TIPS is different from the ones that specialize in Treasuries; this is evident in the discussion found in Fleckenstein et al. (2014), and confirmed by our own conversations with a TIPS trader. Obviously, if every dealer who trades nominal Treasuries had an equal level of expertise in trading TIPS, that would be a step in the right direction (i.e., towards a more consolidated secondary market). However, it is not entirely clear that the Treasury has the power to enforce such a change. (Also, understanding why some dealers who actively trade nominal Treasuries choose to not trade TIPS is beyond the scope of our paper.)

Aside from working towards an integrated dealer network that specializes equally in the trade of TIPS and nominal Treasuries, are there other regulations that can promote secondary market consolidation? Perhaps some useful lessons can be learned from another type of fixed-income OTC secondary markets that have been traditionally known to be segmented, but have recently made positive steps towards a greater degree of integration. In a 2014 report (BlackRock, 2014), leading executives of Blackrock (the world's largest asset manager) characterize the secondary market for corporate bonds as "broken" and "fragmented", and list a number of actions to remedy this fragmentation and increase secondary market liquidity. The list of suggestions includes: i) more "all-to-all" trading venues—not just "dealer-to-customer" or "dealer-to-dealer"; ii) adoption of multiple electronic trading (e-trading) protocols—not just request for quote (RFQ) or central limit order book (CLOB); iii) standardization of selected features of newly-issued corporate bonds; [15] and iv) behavioral changes by market participants recognizing the fundamentally changed landscape.

While, obviously, secondary markets for corporate bonds are a different entity, it

[15] This is the only recommendation on the list that is essentially under the control of the Treasury. The "selected features" that the authors suggest should be standardized include the bond's maturities. Specifically, the authors believe that very diverse offerings of maturities exacerbate market segmentation, and they advocate for a standardization "via steps such as issuing similar amounts and maturities at regular intervals", adding that "the standardized terms would improve the ability to quote and trade bonds". A quick look at the maturities currently offered by Treasuries and TIPS indicates that there is indeed room for further standardization. The Treasury offers nominal Treasuries with maturities of 4, 8, 13, 17, 26, and 52 weeks (Bills), 2, 3, 5, 7, and 10 years (Notes), and 20 and 30 years (Bonds). On the other hand, TIPS currently offer maturities of 5, 10, and 30 years.

seems plausible that some of these suggestions can be applicable for implementing a more consolidated secondary market for Treasury-issued securities, after the necessary adjustments to fit the institutional details of these markets. In the same report, the authors state that "these are not just regulatory changes, but much broader reforms—to fix corporate fixed income markets will require changes in behavior by all market participants—issuers, intermediaries and investors. And yes, by regulators, too." This highlights another similarity between corporate bonds and Treasuries, which we alluded to earlier in this section: it is not obvious how a perfect consolidation of the secondary markets for TIPS and nominal Treasuries can be achieved in the real world. But it is pretty obvious that it requires work from several involved parties, not just the U.S. Treasury. (Perhaps if the consolidation of these markets were entirely in the hands of the U.S. Treasury, it would have already been achieved.)

6 Exogenous Differences: Buy-and-Hold Investors

The analysis so far has established that our model can explain the illiquidity of TIPS, both qualitatively and quantitatively, without the need to assume exogenous differences between the markets where the two assets trade or the investors who hold them. That said, some of these exogenous differences assumed in the literature are realistic and relevant. Thus, one might wonder how the adoption of such an assumption would interact with our amplification mechanism in general equilibrium. To that end, we explore an extension of our model where some agents "*buy and hold*" TIPS, as in Andreasen et al. (2021) and other papers. This idea can be easily embedded in our framework by assuming that some agents purchase bonds in the primary market, but they hold them to maturity, so that they will not participate in any trade in the secondary OTC market (and, importantly, the other agents know/expect that). In that sense, the presence of buy-and-hold investors effectively acts as a *reduction* in the supply of TIPS that is available for trade in OTC$_B$—a reduction which, through the endogenous market entry mechanism of our model, can further exacerbate the illiquidity of TIPS.

More formally, we assume that the new type of investors *buy* a fraction $\Gamma \in (0, 1)$ of the outstanding supply of TIPS in the CM. But since these agents never receive a consumption opportunity in the DM, they have no reason to sell assets in OTC$_B$, and simply *hold* them until they mature in next period's CM. Everything else in the model remains unaltered. Importantly, a measure 1 of 'regular' agents optimally choose to specialize in asset A (Treasuries) or B (TIPS), and, accordingly, attempt to liquidate their assets in

Γ	θ	δ	ρ	Liquidity Premium $L_A = L_B$	Yield		Gain (USD Millions)
					Treasuries	TIPS	
0	0.7	1.8853	0.0092	2.6987%	2.5997%	0.2934%	694.19
0	0.8	1.6392	0.0089	2.6985%	2.5999%	0.2936%	688.26
0	0.9	1.4477	0.0086	2.6979%	2.6005%	0.2942%	669.97
0.1	0.7	1.8795	0.0089	2.7013%	2.5972%	0.2909%	772.73
0.1	0.8	1.6340	0.0086	2.7009%	2.5975%	0.2912%	760.95
0.1	0.9	1.4430	0.0083	2.7004%	2.5980%	0.2917%	745.85
0.2	0.7	1.8736	0.0086	2.7041%	2.5944%	0.2881%	857.51
0.2	0.8	1.6288	0.0083	2.7037%	2.5947%	0.2885%	847.32
0.2	0.9	1.4384	0.0080	2.7033%	2.5951%	0.2889%	833.88
0.4	0.7	1.8620	0.0079	2.7100%	2.5884%	0.2823%	1,037.4
0.4	0.8	1.6185	0.0076	2.7098%	2.5887%	0.2826%	1,030.1
0.4	0.9	1.4292	0.0074	2.7095%	2.5889%	0.2828%	1,022.7

Table 5: Results with buy-and-hold investors. The first column shows Γ, the fraction of the asset B supply held by buy-and-hold investors. The second, third, and fourth columns display combinations of (θ, δ, ρ) that align the model with the observed data in Table 2 for each given Γ. The remaining columns present the counterfactual results for each combination of $(\Gamma, \theta, \delta, \rho)$ in a model with a consolidated secondary asset market.

OTC_A or OTC_B, upon the arrival of the idiosyncratic consumption shock.[16] The regular agents correctly perceive that an amount ΓB of TIPS, absorbed by the buy-and-hold agents, will not be available for trade in OTC_B. Of course, this will be reflected in their estimated surplus from trading in OTC_A versus OTC_B, and the consequent matching probabilities in the two OTC markets, ultimately driving their entry decisions.

The details of the model extension with buy-and-hold agents can be found in Appendix B, and the results are presented in Table 5. Since calculating the fraction of TIPS that are absorbed by buy-and-hold investors in the real world is not easy, we present results for various Γ's, specifically, $\Gamma = 0$ (baseline model) and $\Gamma = 0.1, 0.2, 0.4$, to examine how the varying values of Γ interact with our amplification mechanism in general equilibrium; these values can be seen in the first column of the table. The second, third, and

[16] The buy-and-hold agents are assumed to have a perfectly inelastic demand for (ΓB units of) TIPS. Strictly speaking, this is not fully optimal: since these agents are not subject to the idiosyncratic consumption shock, they should be willing to buy asset B only at its fundamental price, and not pay a liquidity premium. The idea here, borrowed from the related finance literature, is that these agents are a special class, and, for reasons that are outside of the model(s), they must hold *some* asset, e.g., because that is the only way to transfer wealth across periods. Given this, purchasing TIPS, i.e., the less liquid, hence cheaper, of the two assets, *is* optimal. (For the same reasons, buy-and-hold agents never carry money.)

fourth columns display combinations of (θ, δ, ρ) that align the model with the observed data in Table 2 for each given Γ. The remaining columns present the counterfactual results for each combination of $(\Gamma, \theta, \delta, \rho)$ in a model with a consolidated secondary asset market.

The main result is that the addition of buy-and-hold agents *enhances* the quantitative power of our amplification mechanism, and this can be seen in two ways. First, other things equal, increasing values of Γ, i.e., increasing supplies of TIPS held by buy-and-hold investors, imply that the degree of IRS in matching needed to fit the data *drops* even further. This is illustrated by the decreasing values of ρ as one moves downwards from the baseline model ($\Gamma = 0$) to greater values of Γ. Second, for given parameter values, the gain the U.S. Treasury would obtain if the two secondary markets were perfectly consolidated *increases* even further (the last column of Table 5). As seen in Section 5, in the baseline model ($\Gamma = 0$), a perfect consolidation of the two OTC markets would help the U.S. Treasury save between 670 and 694.2 million dollars per year, on average, on the given issuance of Treasuries and TIPS (in their 10-year maturity segment) during the post-Great Recession to pre-Covid period. Incorporating buy-and-hold investors significantly boosts these numbers, so much so that if 40% of the TIPS supply is held by these agents ($\Gamma = 0.4$), the savings of the Treasury from the consolidation of OTC$_A$ and OTC$_B$ would exceed a billion dollars per year, on average (the last three rows of Table 5).

The key intuition is that 'regular' agents realize that whatever supply of asset B is absorbed by buy-and-hold investors *will never make its way to OTC$_B$* for secondary market trade. But it is now well-understood that this lower amount of TIPS available for OTC trade manifests itself as a liquidity advantage for OTC$_A$, which is further magnified by our amplification mechanism (explained in the Introduction and in Section 4), thus intensifying the TIPS illiquidity. Since higher values of Γ effectively mean lower supplies of TIPS available for trade in OTC$_B$, our amplification mechanism becomes stronger as Γ increases, requiring an even lower degree of IRS in matching (ρ) in order to fit the observed yield data. Moreover, since a higher Γ exacerbates the illiquidity of TIPS, and consolidating the two OTC markets remedies that problem, it follows that the gains of the U.S. Treasury from the merging of the two secondary markets are also increasing in Γ.

An interesting observation that follows from the exercise performed in this section is illustrated in the fifth column of Table 5, which reports the liquidity premia of the two assets if the two secondary markets were perfectly consolidated. The careful reader may have observed that in the counterfactual environment of a unique, consolidated OTC market the liquidity premia of Treasuries and TIPS are equalized, *even if Γ is positive* (and regardless of how large it is). This means that the addition of buy-and-hold investors in our model (an ingredient which is admittedly realistic) cannot explain the illiquidity of

TIPS in its own right. As we already claimed in Section 2.1, assuming some segmentation between the secondary markets for TIPS and Treasuries is a *necessary* condition for capturing the well-documented illiquidity of TIPS. At the same time, the addition of buy-and-hold investors makes the message of our paper even stronger, as it highlights that the U.S. Treasury leaves substantial profits on the table by having Treasuries and TIPS trade in segmented secondary markets.

7 Conclusion

This paper aims to rationalize, qualitatively and quantitatively, the well-documented illiquidity of Treasury Inflation-Protected Securities (TIPS) vis-à-vis nominal Treasuries, which has been characterized as puzzling, since one would expect that TIPS should also enjoy the "convenience yield" of U.S. Treasury-issued securities. In our model, agents who have received an idiosyncratic consumption shock can sell bonds for money to boost their liquidity in OTC secondary asset markets. Treasuries and TIPS trade in distinct secondary markets, and agents visit the market where they expect to find the best terms.

Our mechanism can explain the superior liquidity of Treasuries exploiting only one difference between Treasuries and TIPS that is perfectly observed in the data: the far greater supply of Treasuries. Since Treasuries come in larger supply, more such assets are likely to be offered for sale in their respective secondary market, and thus potential meetings in that market will generate a larger trading surplus. This advantage originally attracts more buyers into the secondary market for Treasuries, which implies a higher trading probability for potential sellers 'down the road', and manifests itself as a greater ex-ante willingness to pay a higher liquidity premium for Treasuries. With even minimal increasing returns to scale in matching, our amplification mechanism can be so powerful that asset demand curves become upward sloping, so that the asset in larger supply—here, the Treasuries—emerges endogenously as the more liquid asset.

A calibrated version of our model can match the observed yields and liquidity premia of TIPS and Treasuries. We use our model to carry out a counterfactual exercise that contributes to the recent debate about whether the U.S. Treasury should phase out TIPS, and how much money they are leaving on the table by issuing securities that are not as highly valued by the market as nominal Treasuries. Our model suggests that the Treasury does not need to eliminate TIPS; instead, it should seek to reduce segmentation and ensure that TIPS trade alongside nominal Treasuries in consolidated secondary markets. Specifically, we find that the U.S. Treasury would save up to 700 million dollars per year on the issuance of 10-year Treasuries and TIPS during the post-Great Recession to pre-

Covid period, if the two assets were traded in a consolidated secondary market.

Although our model can explain the illiquidity of TIPS vis-à-vis nominal Treasuries without the *need* to assume exogenous differences between the investors who hold the two assets, we explore an extension with investors who "buy and hold" TIPS (an assumption commonly adopted in related literature). The presence of buy-and-hold investors effectively acts as a reduction in the supply of TIPS that is available for trade in the OTC market, thus further exacerbating the illiquidity of TIPS. In sum, adding buy-and-hold investors both enhances our quantitative results and reinforces our theoretical conclusion that the observed liquidity differentials between the two asset types point to a significant degree of secondary market segmentation.

Appendix

A Consolidated Secondary Asset Market

Matching probabilities The measures of asset sellers and buyers in the consolidated OTC market are ℓ and $1-\ell$, respectively. The matching probabilities for agents of type $i \in \{C, N\}$, denoted by α_i, are then given by

$$\alpha_C = \frac{f(\ell, 1-\ell)}{\ell} = \delta(1-\ell), \quad \alpha_N = \frac{f(\ell, 1-\ell)}{1-\ell} = \delta\ell.$$

Value functions The CM value function for a consumer entering the market with m units of money and d_j units of asset $j \in \{A, B\}$ is given by

$$\begin{aligned} W(m, d_A, d_B) = & \max_{\substack{X,H,\\ \hat{m},\hat{d}_A,\hat{d}_B}} \left\{ X - H + \beta\, \mathbb{E}_i\big[\Omega_i(\hat{m}, \hat{d}_A, \hat{d}_B)\big] \right\} \\ \text{s.t. } & X + \varphi(\hat{m} + p_A\hat{d}_A + p_B\hat{d}_B) = H + \varphi(m + d_A + (1+\pi)d_B + \mu M), \end{aligned}$$

where Ω_i represents the value function for a consumer of type $i \in \{C, N\}$ entering the consolidated OTC market. The expectation operator $\mathbb{E}_i$ is with respect to the idiosyncratic consumption shock, and $\mathbb{E}_i[\Omega_i] = \ell\,\Omega_C + (1-\ell)\Omega_N$, reflecting that the representative buyer becomes a C-type with probability ℓ.

The OTC value function for a type-i consumer is given by

$$\begin{aligned} \Omega_C(m, d_A, d_B) &= \alpha_C V\,(m+\xi, d_A - \chi_A, d_B - \chi_B) + (1-\alpha_C)\, V(m, d_A, d_B), \\ \Omega_N(m, d_A, d_B) &= \alpha_N W\,(m-\xi, d_A + \chi_A, d_B + \chi_B) + (1-\alpha_N)\, W(m, d_A, d_B), \end{aligned}$$

where ξ is the amount of money transferred to the C-type, χ_j the amount of asset j transferred to the N-type, and V the consumer's value function in the DM.

Denoting q as the quantity of special goods traded and τ as the total payment in money, the DM value function for a consumer is given by

$$V(m, d_A, d_B) = u(q) + W(m-\tau, d_A, d_B).$$

The producer's value functions are unchanged from the baseline model.

Terms of trade The terms of DM trade remain as in the baseline model, given by (4) and (5). For the consolidated OTC market, consider a meeting between a C-type consumer with portfolio (m, d_A, d_B) and an N-type consumer with portfolio $(\widetilde{m}, \widetilde{d}_A, \widetilde{d}_B)$. The two parties negotiate a transfer of money, ξ, to the C-type and a combination of assets A and B, (χ_A, χ_B), to the N-type. The surplus is split via proportional bargaining, with $\theta \in (0, 1)$ denoting the C-type's bargaining power, subject to the C-type's asset holding constraints and the N-type's money holding constraint. The surpluses for both agents are given by

$$\begin{aligned} S_C &= V(m + \xi, d_A - \chi_A, d_B - \chi_B) - V(m, d_A, d_B), \\ S_N &= W(\widetilde{m} - \xi, \widetilde{d}_A + \chi_A, \widetilde{d}_B + \chi_B) - W(\widetilde{m}, \widetilde{d}_A, \widetilde{d}_B). \end{aligned}$$

Due to the linearity of the value function W, $S_C = u(q(m+\xi)) - u(q(m)) - \varphi\chi_A - \varphi(1+\pi)\chi_B$ and $S_N = -\varphi\xi + \varphi\chi_A + \varphi(1 + \pi)\chi_B$. The bargaining problem is then described by

$$\max_{\xi, \chi_A, \chi_B} S_C \quad \text{s.t.} \quad S_C = \frac{\theta}{1-\theta} S_N, \quad \chi_A \leq d_A, \quad \chi_B \leq d_B, \quad \text{and} \quad \xi \leq \widetilde{m}.$$

The Kalai constraint implies

$$\varphi\chi_A + \varphi(1 + \pi)\chi_B = (1 - \theta)(u(q(m + \xi)) - u(q(m)) + \theta\varphi\xi \equiv \sigma(\xi),$$

meaning that the asset seller can acquire $\sigma^{-1}(\varphi\chi_A + \varphi(1+\pi)\chi_B)$ units of money by selling χ_A units of asset A and χ_B units of asset B. Since carrying money is costly, a consumer will never bring more than m^* from the CM and will want to acquire the amount needed to reach m^*, namely, $m^* - m$. If she holds enough assets to acquire $m^* - m$, i.e., $\sigma^{-1}(\varphi d_A + \varphi(1 + \pi)d_B) \geq m - m^*$, then she acquires $m - m^*$; otherwise, she sells all her assets and acquire $\sigma^{-1}(\varphi d_A + \varphi(1 + \pi)d_B)$ units of money.[17] The bargaining solution is thus given by

$$\xi(m, d_A, d_B) = \min\{m^* - m, \sigma^{-1}(\varphi d_A + \varphi(1 + \pi)d_B)\}, \tag{A.1}$$

$$(\chi_A(m, d_A, d_B), \chi_B(m, d_A, d_B)) = \begin{Bmatrix} (\widetilde{\chi}_A, \widetilde{\chi}_B) : \\ \varphi\widetilde{\chi}_A + \varphi(1 + \pi)\widetilde{\chi}_B = \sigma(\xi(m, d_A, d_B)), \\ \widetilde{\chi}_A \leq d_A, \ \widetilde{\chi}_B \leq d_B \end{Bmatrix}. \tag{A.2}$$

Optimal portfolio choice Agents choose their optimal portfolio in the CM independently of their prior trading histories. Following the same procedure as in the baseline

[17] As in the baseline model, we assume the N-type's money holdings do not constrain the trade, effectively ignoring the constraint $\xi \leq \widetilde{m}$ in the bargaining problem. The same sufficient condition from the baseline model applies. For more detail, see footnote 7.

model, the objective function is derived as

$$J(\hat{m}, \hat{d}_A, \hat{d}_B) = -\varphi(\hat{m} + p_A \hat{d}_A + p_B \hat{d}_B) + \beta\hat{\varphi}(\hat{m} + \hat{d}_A + (1+\pi)\hat{d}_B) \\ + \beta\ell(u(\hat{\varphi}\hat{m}) - \hat{\varphi}\hat{m} + \alpha_C S_C).$$

The interpretation remains the same as in the baseline model, except that the secondary asset market is now consolidated, and agents no longer need to make an entry choice.

Equilibrium A steady-state equilibrium consists of the equilibrium prices $\{\varphi, p_A, p_B\}$ and quantities $\{q_0, q_1\}$. Here, q_k represents the amount of special goods purchased by a C-type agent in the DM, where $k \in \{0, 1\}$ indicates whether the agent traded ($k = 1$) or did not trade ($k = 0$) in the preceding OTC market.

Once the quantities are determined, the equilibrium prices follow directly. The price of money comes from the money market clearing condition:

$$\varphi M = q_0, \tag{A.3}$$

which states that q_0 equals real balances, given the consumer's TIOLI offers in the DM. The optimal asset portfolio choice gives the asset prices:

$$p_A = \frac{1}{1+i}(1 + L_A), \quad p_B = \frac{1+\pi}{1+i}(1 + L_B), \tag{A.4}$$

where the liquidity premium of asset j, denoted L_j, is given by

$$L_A = L_B = \ell\alpha_C \frac{\theta}{w(q_1)}(u'(q_1) - 1), \tag{A.5}$$

with $w(q_1) \equiv 1 + (1-\theta)(u'(q_1) - 1)$ as before. Note that, unlike in the baseline model, both assets possess the same liquidity premium because, in a consolidated market, they become perfect substitutes in the provision of liquidity.

The quantities satisfy two equilibrium conditions. First, the money demand equation is determined by the optimal money holding decision:

$$i = \ell\left(1 - \alpha_C \frac{\theta}{w(q_1)}\right)(u'(q_0) - 1) + \ell\alpha_C \frac{\theta}{w(q_1)}(u'(q_1) - 1). \tag{A.6}$$

Second, the OTC trading protocol $q_1 = q_0 + \varphi\xi$, combined with the bargaining solution (as in (A.1) and (A.2)) and the market clearing conditions ($A = d_A$, $B = d_B$, $\varphi M = q_0$), gives

$$q_1 = \min\left\{q^*,\, q_0 + \frac{1}{\theta}\frac{A + (1+\pi)B}{M}\,q_0 - \frac{1-\theta}{\theta}(u(q_1) - u(q_0))\right\}. \tag{A.7}$$

A steady-state equilibrium of the model is defined as follows:

Definition 2. For given asset supplies $(A, B) \in \mathbb{R}^2_+$, a steady-state equilibrium is a list $\{\varphi, p_A, p_B, q_0, q_1\}$ that solves (A.3), (A.4), (A.6), and (A.7).

B Models with Buy-and-Hold Investors

Buy-and-hold investors buy and hold a fraction $\Gamma \in (0,1)$ of the supply of asset B for non-liquidity reasons outside the model. These agents do not consume in the DM, do not hold money, and therefore do not participate in OTC trade, making the Γ fraction of asset B unavailable for OTC trade. The remaining 'regular' agents, with a total measure normalized to 1, correspond to the standard agents in the models without buy-and-hold agents. Here, we highlight only the changes introduced by these new agents.

Segmented secondary asset market With the buy-and-hold agents, the market clearing condition for asset B becomes $B = (1 - e_C)d_B + \Gamma B$, and the OTC$_B$ trading protocol adjusts to

$$q_{1B} = \min\left\{q^*,\, q_{0B} + \frac{1}{\theta}\frac{(1+\pi)(1-\Gamma)B}{M}\frac{e_C q_{0B} + (1-e_C)q_{0B}}{1-e_C} - \frac{1-\theta}{\theta}(u(q_{1B}) - u(q_{0B}))\right\},$$

since only the remaining $1 - \Gamma$ fraction of the asset B supply flows into OTC$_B$. All other equilibrium conditions remain unchanged.

Consolidated secondary asset market Similarly, the market clearing condition for asset B becomes $B = d_B + \Gamma B$, and the OTC trading protocol adjusts to

$$q_1 = \min\left\{q^*,\, q_0 + \frac{1}{\theta}\frac{A + (1+\pi)(1-\Gamma)B}{M}\,q_0 - \frac{1-\theta}{\theta}(u(q_1) - u(q_0))\right\}.$$

All other equilibrium conditions remain unchanged.

References

Abrahams, M., T. Adrian, R. K. Crump, E. Moench, and R. Yu (2016). Decomposing real and nominal yield curves. *Journal of Monetary Economics 84*, 182–200.

Afonso, G. and R. Lagos (2015). Trade dynamics in the market for federal funds. *Econometrica 83*(1), 263–313.

Altermatt, L., K. Iwasaki, and R. Wright (2023). General equilibrium with multiple liquid assets. *Review of Economic Dynamics 51*, 267–291.

Andreasen, M. M., J. H. Christensen, and S. Riddell (2021). The TIPS liquidity premium. *Review of Finance 25*(6), 1639–1675.

Berentsen, A., S. Huber, and A. Marchesiani (2014). Degreasing the wheels of finance. *International Economic Review 55*(3), 735–763.

Bethune, Z., M. Choi, and R. Wright (2020). Frictional goods markets: Theory and applications. *Review of Economic Studies 87*(2), 691–720.

BlackRock (2014). Corporate bond market structure: The time for reform is now. Technical report.

Campbell, J. Y. and R. J. Shiller (1996). A scorecard for indexed government debt. *NBER macroeconomics annual 11*, 155–197.

Campbell, J. Y., R. J. Shiller, and L. M. Viceira (2009). Understanding inflation-indexed bond markets. *Brookings Papers on Economic Activity* (Spring 2009), 79–120.

Caramp, N. and S. R. Singh (2023). Bond premium cyclicality and liquidity traps. *Review of Economic Studies 90*(6), 2822–2879.

Chang, B. and S. Zhang (2015). Endogenous market making and network formation. Available at SSRN 2600242.

D'Amico, S., D. H. Kim, and M. Wei (2018). Tips from TIPS: The informational content of Treasury inflation-protected security prices. *Journal of Financial and Quantitative Analysis 53*(1), 395–436.

Dudley, W., J. E. Roush, and M. S. Ezer (2009). The case for TIPS: An examination of the costs and benefits. *Economic Policy Review 15*(1), 1–17.

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

Fleckenstein, M., F. A. Longstaff, and H. Lustig (2014). The TIPS-Treasury bond puzzle. *Journal of Finance 69*(5), 2151–2197.

Fleming, M. J. and N. Krishnan (2012). The microstructure of the TIPS market. *Economic Policy Review 18*, 27–45.

Geromichalos, A. and L. Herrenbrueck (2016). Monetary policy, asset prices, and liquidity in over-the-counter markets. *Journal of Money, Credit, and Banking 48*(1), 35–79.

Geromichalos, A. and L. Herrenbrueck (2022). The liquidity-augmented model of macroeconomic aggregates. *Review of Economic Dynamics 45*, 134–167.

Geromichalos, A., L. Herrenbrueck, and S. Lee (2023a). Asset safety versus asset liquidity. *Journal of Political Economy 131*(5), 1172–1212.

Geromichalos, A., L. Herrenbrueck, and S. Lee (2023b). The strategic determination of the supply of liquid assets. *Review of Economic Dynamics 49*, 1–36.

Geromichalos, A., L. Herrenbrueck, and K. Salyer (2016). A search-theoretic model of the term premium. *Theoretical Economics 11*(3), 897–935.

Geromichalos, A. and I. Simonovska (2014). Asset liquidity and international portfolio choice. *Journal of Economic Theory 151*, 342–380.

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.

Herrenbrueck, L. (2019). Frictional asset markets and the liquidity channel of monetary policy. *Journal of Economic Theory 181*, 82–120.

Herrenbrueck, L. and A. Geromichalos (2017). A tractable model of indirect asset liquidity. *Journal of Economic Theory 168*, 252 – 260.

Herrenbrueck, L. and Z. Wang (2023). Interest rates, moneyness, and the fisher equation. Simon Fraser University Working Paper.

Kalai, E. (1977). Proportional solutions to bargaining situations: Interpersonal utility comparisons. *Econometrica 45*(7), 1623–30.

Kim, D., C. Walsh, and M. Wei (2019). Tips from TIPS: Update and discussions. *FEDS Notes*. Washington: Board of Governors of the Federal Reserve System.

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

Lagos, R. (2010). Asset prices and liquidity in an exchange economy. *Journal of Monetary Economics 57*(8), 913–930.

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

Lagos, R., G. Rocheteau, and P.-O. Weill (2011). Crises and liquidity in over-the-counter markets. *Journal of Economic Theory 146*(6), 2169–2205.

Lagos, R. and R. Wright (2005). A unified framework for monetary theory and policy analysis. *Journal of Political Economy 113*(3), 463–484.

Lee, S. and K. M. Jung (2020). A liquidity-based resolution of the uncovered interest parity puzzle. *Journal of Money, Credit and Banking 52*(6), 1397–1433.

Mattesini, F. and E. Nosal (2016). Liquidity and asset prices in a monetary model with OTC asset markets. *Journal of Economic Theory 164*, 187–217.

Pflueger, C. E. and L. M. Viceira (2016). Return predictability in the Treasury market: Real rates, inflation, and liquidity. In P. Veronesi (Ed.), *Handbook of Fixed-Income Securities*, Chapter 10, pp. 191–209. John Wiley & Sons, Ltd.

Rocheteau, G., R. Wright, and C. Zhang (2018). Corporate finance and monetary policy. *American Economic Review 108*(4–5), 1147–1186.

Sack, B. P. and R. Elsasser (2004). Treasury inflation-indexed debt: A review of the U.S. experience. *Economic Policy Review 10*(1), 47–63.

Üslü, S. (2019). Pricing and liquidity in decentralized asset markets. *Econometrica 87*(6), 2079–2140.

Vayanos, D. and T. Wang (2007). Search and endogenous concentration of liquidity in asset markets. *Journal of Economic Theory 136*(1), 66–104.

Vayanos, D. and P.-O. Weill (2008). A search-based theory of the on-the-run phenomenon. *Journal of Finance 63*(3), 1361–1398.

Weill, P.-O. (2007). Leaning against the wind. *Review of Economic Studies 74*(4), 1329–1354.

Weill, P.-O. (2008). Liquidity premia in dynamic bargaining markets. *Journal of Economic Theory 140*(1), 66–96.