Intermediary market power and capital constraints*

Jason Allen+ and Milena Wittwer§

July 15, 2023

Abstract

We examine how intermediary capitalization affects asset prices in a framework that allows for intermediary market power. We introduce a model in which capital constrained intermediaries buy or trade an asset in an imperfectly competitive market, and show that weaker capital constraints lead to both higher prices and intermediary markups. In exchange markets, this results in reduced market liquidity, while in primary markets, it leads to higher auction revenues at an implicit cost of larger price distortion. Using data from Canadian Treasury auctions, we demonstrate how our framework can quantify these effects by linking asset demand to individual intermediaries' balance sheet information.

Keywords: Financial intermediaries, market power, price impact, asset demand, asset pricing, government bonds, Basel III, capital requirements, leverage ratios

JEL: G12, G18, G20, D40, D44, L10

*Previous versions of the paper were circulated under the title "Intermediary asset pricing: Capital constraints and market power". The views presented are those of the authors and not necessarily those of the Bank of Canada. Any errors are our own. We thank the Triton team at the Bank of Canada, Rui Albuquerque, Eric Budish, David Cimon, Robert Clark, Darrell Duffie, Amy Huber, Mark Egan, Mahyar Kargar, Leonid Kogan, Liran Einav, Matthew Gentzkow, Daniel Smith, Or Shachar, Monika Piazzesi, Marzena Rostek, Fabricius Somogyi, Xavier Vives, Ji Hee Yoon, in addition to all participants of seminars at Boston College, University of Maryland, Stanford, Boston Conference on Markets and Competition, European Winter Meeting of the Econometric Society, NBER Market Design, NYU IO day and the Annual Meeting of Women in Microstructure, SFS Cavalcade, FIRS, EFA. Correspondence to: +Jason Allen (Bank of Canada): jallen@bankofcanada.ca and §Milena Wittwer - Boston College, Email: wittwer@bc.edu.

1 Introduction

What moves asset prices is one of the oldest questions in finance. The intermediary asset pricing literature suggests that the prices of many assets depend not only on the preferences of households, but also on the equity capitalization of financial intermediaries, called dealers (e.g., Gromb and Vayanos (2002); Brunnermeier and Pedersen (2009); He and Krishnamurthy (2012, 2013); Brunnermeier and Sannikov (2014)). In this literature, dealers typically face funding or capital constraints and execute trades in perfectly competitive markets. In practice, however, dealers enjoy market power—as documented for various trade settings, including Treasury, repo, foreign exchange, mortgage backed securities, and equity securities lending markets (e.g., Wallen (2022); Allen and Wittwer (2023); An and Song (2023); Chen et al. (2023); Huber (2023)).

Our contribution is to study how dealer capitalization affects asset prices and markups, and quantify the effect, in a framework that allows for dealer market power (as in Wilson (1979); Klemperer and Meyer (1989); Kyle (1989); Vives (2011); Rostek and Weretka (2012); among others). We introduce a model in which capital constrained dealers buy (or trade) assets in an imperfectly competitive market, and estimate it with data on Canadian Treasury auctions.[1]

In the model, presented in Section 2, risk averse dealers compete to buy (or trade) multiple units of an asset that pays out an uncertain return in the future. They are subject to a capital constraint, which depends on the auction outcome, and may have private information about their own balance sheet. In addition, they may be uncertain about the auction supply. The market clears via one out of two auction formats, which represent different financial markets, including primary auctions and exchanges. In the benchmark model, dealers submit decreasing demand functions that specify how much they are willing to pay for different units

[1] Capital requirements aim to strengthen the risk management of banks and avoid the build up of systemic risks. Our analysis does not incorporate how these risks change when relaxing constraints.

of the asset; the market clears at the price at which aggregate dealer demand meets supply, and each dealer wins the amount it asked for at that price (uniform price auction). In the extended model, winning dealers pay the prices they bid (discriminatory price auction).

Solving for an equilibrium is challenging, because point-wise maximization—a common approach in the literature—does not work when bidders face outcome-dependent constraints. Instead, we must consider all feasible demand functions. By doing so, we derive necessary conditions for symmetric Bayesian Nash Equilibria (hereinafter referred to as equilibria). Moreover, we establish that there is no linear equilibrium when bidders have private information, but derive a unique symmetric linear equilibrium for auctions in which bidders face common uncertainty about supply.

Our model highlights two effects of relaxing capital constraints. On the one hand, the market price increases. This is due to the fact that as the shadow costs of the capital constraint decrease, it becomes cheaper for dealers to purchase larger quantities of the asset. On the other hand, dealers exert greater influence on the market price, deviating it further from the price that would result if the market was perfectly competitive. This means that the price distortion due to market power increases, which reduces market liquidity in exchange markets. The effect is absent in models with perfect competition, and intuitively stems from the increased flexibility of dealers to manipulate market outcomes to their advantage when they face fewer constraints.

To quantify these effects, we use data on Canadian Treasury auctions, presented in Section 3, leveraging two attractive features. First, dealers submit entire demand curves. We can therefore observe whether demand is flat or steep, which is the main mechanism through which shadow costs affect prices in our model. Alternatively, we would need to aggregate individual demands from secondary market trades. Second, we can link the dealers' demand curves to balance sheet information, which is crucial for establishing a connection between dealer demand and capitalization.

Our data combine bidding information on all Canadian government bond auctions between January 2019 and February 2022 with balance sheet information of the eight largest dealers (at the company holding level following He et al. (2017)), and trade-level information from the secondary market. We observe all winning and losing bids, and can identify each bidder thanks to unique identifiers. In addition, we see the Basel III Leverage Ratio (LR) of each dealer, which is the Canadian equivalent to the Supplementary Leverage Ratio (SLR) in the U.S. It is reported quarterly, measures a bank's Tier 1 capital relative to its total leverage exposure, and must be above a regulatory threshold, which we also observe. Lastly, we gather data on all secondary market trades conducted by dealers to observe the volatility of the returns they obtain from purchasing bonds at auction and subsequently selling them in the secondary market. To make bonds with different maturities and coupon payments more comparable, we express all empirical findings using yields-to-maturity rather than prices.

In Section 4, we estimate the two key parameters of the model: the dealer's risk aversion and the shadow costs of the capital constraint. To accomplish this, we employ estimation techniques from the auctions literature (introduced by Guerre et al. (2000); Hortaçsu and McAdams (2010); Kastl (2011)) to estimate each bidder's willingness to pay at a discrete number of points. Then we fit the model-implied functional form for the willingness to pay through these points. Finally, we take advantage of a temporary exemption of domestic government bonds from the LR during the COVID-19 pandemic to identify the degree of dealer risk aversion and their shadow costs of the capital constraint by analyzing how the willingness to pay varies around the policy change.

We find that dealers are moderately risk-averse and face sizable shadow costs of the capital constraint. In fact, the median cost (of 3.5%) is as high as the typical markup a dealer charges its clients (i.e., the median difference between the price at which a dealer buys a bond at auction and the price at which she sells this bond in the secondary market). This suggests that dealers barely break even, and might explain why so many dealers have left the market (as documented by Allen et al. (2023)).

A back-of-the-envelop calculation tells us that the market yield decreases and the yield distortion due to bid shading increases by 3.4 bps when the shadow cost of capital decreases by 1%. This highlights that relaxing capital constraints leads to a reduction in bond yields, which overall increases auction revenues, at an implicit cost of larger yield distortion due to market power. When the interest rate level is high, these effects are economically meaningful. In our sample period, where rates are low, however, the effects are small. This suggests that the Canadian regulator did not face a quantitatively meaningful trade-off when deciding whether to relax or tighten capital constraints during the COVID-19 pandemic.

To conclude our study, in Section 5 we draw a closer connection to the intermediary asset pricing literature by extending our analysis to study how intermediary market power affects whether commonly considered intermediary frictions (such as moral hazard or capital constraints) matter for asset prices. We show that the price effect of these frictions depends on the degree of market power. Hopefully, this motivates future research that can analyze the implications of intermediary market power in a macroeconomic model of intermediary asset pricing, and empirical research to assess the degree of competition in different financial markets.

Related literature. We contribute to five distinct strands of the literature.

The paper's topic fits into an ample intermediary asset pricing literature that examines the impact of dealer capitalization (or leverage) on asset price behavior due to constraints on debt (e.g., Brunnermeier and Pedersen (2009); Pedersen and Gârleanu (2011); Adrian and Shin (2014); Moreira and Savov (2017); Elenev et al. (2021)), or constraints on equity (e.g., He and Krishnamurthy (2013, 2012); Brunnermeier and Sannikov (2014)). Given our focus on banks, we follow He et al. (2017) and rely on equity constraints. The key difference relative to these (macroeconomic) models is that we zoom in on the market in which intermediaries buy or trade assets, and allow dealers to impact prices as a result of market power in the tradition of Kyle (1989).[2]

[2]Our extended model, presented in Section 5, is more similar to the intermediary asset pric-

The market clears via a multi-unit auction following Wilson (1979), Kyle (1989), and Klemperer and Meyer (1989). More recent contributions include Vayanos (1999); Vives (2011); Gârleanu and Pedersen (2013); Rostek and Weretka (2012, 2015); Malamud and Rostek (2017); Du and Zhu (2017); Kyle et al. (2017); Bergemann et al. (2021); Wittwer (2021); Rostek and Yoon (2021); Zhang (2022). Our innovation in this literature is introducing bidder constraints that are dependent on the auction outcome, such as capital constraints (which can be generalized to other constraints such as budget constraints). The presence of such constraints implies that common tools of the literature are not applicable. For instance, it is no longer possible to solve for a linear equilibrium by point-wise maximization.

Our empirical analysis adds to a growing literature on the relation between intermediary costs or constraints and asset prices (e.g., Adrian and Shin (2010); Ang et al. (2011); Adrian et al. (2014); He et al. (2017, 2022); Du et al. (2018); Check et al. (2019); Gospodinov and Robotti (2021); Kargar (2021); Haddad and Muir (2021); Baron and Muir (2022); Fontaine et al. (2022); Du et al. (2023a,b); Huang et al. (2023)). Most existing studies use market-level data, such as cross-sectional returns of different asset classes, and rely on proxy variables to capture intermediary costs, such as the VIX, or aggregate capital holdings. We zoom in on one market in which we can link dealer demand with balance sheet information to establish a direct relationship between dealer capitalization and asset demand. Further, we estimate two important parameters: the shadow costs of the capital constraint and the dealer's degree of risk aversion. For this, we construct our own volatility measure using secondary market trade data.

---

ing literature, which abstracts from market power with few recent exceptions (e.g. Corbae and D'Erasmo (2021); Jamilov (2021); Villa (2022); Wang et al. (2022)). These papers introduce monopolistic or oligopolistic (Cournot) competition of banks vis-a-vis firms or consumers, while we analyze market power in a trade setting. Therefore, our insights, especially those on the linkage between intermediary market power and capital constraints, are fundamentally different from those found in this literature. For instance, Villa (2022) finds that banks exert more market power on firms that are more financially constrained. In our setting, banks, i.e., dealers, exert less market power when constraints tighten.

For estimation, we adopt techniques from the literature on multi-unit auctions, developed by Guerre et al. (2000), Hortaçsu and McAdams (2010) and Kastl (2011) and extended by Hortaçsu and Kastl (2012), and Allen et al. (2020, 2023). This literature commonly assumes that financial institutions are risk-neutral; with the exception of Gupta and Lamba (2017), who exogenously choose a risk aversion parameter to simulate their model. However, the assumption of risk-neutrality stands in contrast to the related market microstructure literature which builds on Kyle (1989), and assumes that financial institutions have preferences with constant absolute risk aversion (CARA). We follow this literature and impose CARA preferences to circumvent the impossibility result by Guerre et al. (2009) that one cannot non-parametrically identify risk aversion (in first-price auctions).

This approach is similar to a handful of papers that estimate risk-aversion in auctions for procurement, timber, and other non-financial goods (e.g., Campo et al. (2011); Campo (2012); Bolotnyy and Vasserman (2023); Häfner (2023); Luo and Takahashi (2023)). It complements the common macroeconomic practice of calibrating risk aversion for households using Euler equations (e.g., Attanasio and Weber (1989); Vissing-Jørgensen (2002)). Since the risk aversion of intermediaries plays a crucial role in intermediary asset pricing models, our estimate can provide valuable input for calibrating these models.

Convention. Throughout the paper, we denote random variables in **bold**, and refer to the markup as the difference between the price at which the market would clear if it was perfectly competitive and the price at which it clears under imperfect competition; or equivalently, as the difference between the yield at which the market clears under imperfect competition versus perfect competition. In a uniform price auction, the markup increases in price impact—a common object of interest.

2 Model

Our goal is to study how both the market price and the markup change when capital constraints are relaxed and dealers have market power. In our benchmark, we model market-clearing via a uniform price auction, in which winning bidders pay the market clearing price. In Appendix A we derive analogous results for discriminatory price auctions, in which winning bidders pay their own bids. The market may be one-sided, meaning that bidders buy but not sell, or double-sided, so that bidders buy and sell. In order to facilitate the comparison with the empirical analysis, we present our framework using a one-sided market, but explain how to adjust it to represent a double-sided market.

In practice, some primary markets, for instance in the U.S., clear via one-sided uniform price auctions, while others, for instance in Canada, clear via one-sided discriminatory price auctions. Trading on an exchange can be approximated via a double-sided uniform price auction, where packages of limit orders form demand schedules (e.g., Kyle (1989)).[3]

Proofs are in Appendix E.

2.1 Players, preferences, and constraints

There are $N > 2$ dealers who compete for units of an asset in an auction. When there are finitely many dealers, each one has some market power in that it can impact the market clearing price. When $N \to \infty$ each dealer is a price-taker, and the market is perfectly competitive.

Total supply $\boldsymbol{A}$ is random; it is drawn from some continuous distribution with support $[0, \overline{A}]$ and has a strictly positive density.[4] In our empirical application, supply is random

[3]The exchange market we model is populated by dealers (who are strategic traders that face capital constraints) and noise-traders. We abstract from strategic investors (who don't face capital constraints). This reflects the fact that it is common for non-dealers to invest via dealers (or brokers) on exchanges, rather than directly.

[4]The support could equivalently be $[0, \infty)$, but all proofs would need to be adjusted slightly.

because dealers don't know the issuance size when they compete. In other settings, the supply might be random due to noise traders.

One unit of the asset pays a return of $\boldsymbol{R} \sim N(\mu, \sigma)$ in the future. In our empirical application, where the asset is a government bond, R represents the price obtained from selling the bond post-auction, which is unknown at the time of the auction.

Before bidding, each dealer draws a multi-dimensional signal, $\boldsymbol{\theta_i}$. This signal includes information about the balance sheet: dealer i has an existing inventory $\boldsymbol{z_i}$ of the asset and holds $\boldsymbol{E_i}$ in equity capital—we normalize the rest of each dealer's balance sheet to zero. The signal is either the private information of dealer i or commonly known by all dealers. When the signal is private, it is drawn independently across dealers from some continuous distribution on bounded support and strictly positive density. In this case, dealers face private and aggregate uncertainty when bidding. When the signal is observed by all dealers, we assume that all dealers are identical: $E_i = E, z_i = z$. We introduce the framework for the more general case with private information. Without private information, we simply omit θ_i in all expressions.

Given signal θ_i, each dealer submits a decreasing (inverse) demand schedule: $p_i(\cdot, \theta_i) : \mathbb{R}^+ \to \mathbb{R}^+$, which specifies how many units of the asset, a, the dealer seeks to buy for price, $p_i(\cdot, \theta_i)$. We denote it's inverse by $a_i(p, \theta_i) = p^{-1}(p, \theta_i)$, if it exists. In a double-sided market, such as an exchange, the demand schedule represents demand net of supply.

To develop the theory, we assume that demand functions are twice continuous and strictly decreasing, and denote the set of functions with that property by $\mathcal{B}$. Working with continuous demand functions is common in the related theory literature in order to achieve tractability, even though in practice demand functions are often discrete. For example, bidders must submit step functions in most Treasury auctions. Therefore, we also provide equilibrium conditions for step-functions in Appendix A.

Once all dealers have submitted their demand curves, the auction clears at the price, P^c, such that aggregate demand meets total supply:

$$P^c : \sum_i a_i(P^c, \theta_i) = A. \tag{1}$$

Each dealer pays the market clearing price, $P^c = p_i(a_i^c, \theta_i)$, for the amount won, $a_i^c = a_i(P^c, \theta_i)$ at that price. To highlight equilibria, we refer to the equilibrium market clearing price by P^* and the winning amount by a_i^*.

Each dealer chooses their demand function to maximize their expected CARA utility from earning wealth, $\omega_i(\boldsymbol{a_i^c}, \boldsymbol{P^c})$, that is generated at market clearing:

$$U(p_i(\cdot, \theta_i)) = \mathbb{E}\left[1 - \exp\left(- \rho\omega_i(\boldsymbol{a_i^c}, \boldsymbol{P^c})\right) \middle| \theta_i\right]. \tag{2}$$

Parameter $\rho > 0$ measures the dealer's degree of risk-aversion.[5] Future wealth, $\omega_i(\boldsymbol{a_i^c}, \boldsymbol{P^c})$, is equal to the asset payoff, $\boldsymbol{R}$, net of the price paid,

$$\omega_i(\boldsymbol{a_i^c}, \boldsymbol{P^c}) = [\boldsymbol{a_i^c} + z_i]\boldsymbol{R} - \boldsymbol{P^c}\boldsymbol{a_i^c}. \tag{3}$$

Motivated by the Basel III requirement that states banks must hold sufficient equity capital, E_i, relative to its total balance sheet exposure, $\boldsymbol{P^c}[\boldsymbol{a_i^c}+z_i]$, each dealer faces capital constraint,

$$\kappa\mathbb{E}[\boldsymbol{P^c}[\boldsymbol{a_i^c} + z_i]|\theta_i] \leq E_i, \text{ where } \kappa > 0. \tag{4}$$

For example, according to Basel III, κ is 3%. We denote the Lagrange multiplier of this constraint by $\lambda_i \geq 0$.

When bidding, the dealer does not yet know where the auction will clear, and therefore

[5]A common alternative to CARA preferences are preferences with constant relative risk aversion (CRRA) (e.g., Bajari and Hortaçsu (2005); Vissing-Jørgensen (2002); He and Krishnamurthy (2013); Gupta and Lamba (2017)). With uniform price auctions, this alternative is intractable.

takes an expectation of the capital constraint. This timing assumption is motivated by the fact that capital requirements must be reported at the end of a quarter, rather than on a daily basis, and that dealers anticipate that they must provide evidence that they were holding sufficient capital, on average, over the course of the quarter.

As an alternative to the capital constraint, we could assume that each dealer faces a balance sheet cost that depends on the nominal amount of the asset that the dealer holds on her balance sheet post-auction. One simple functional form for such a balance sheet cost is $mc_i P^c[a_i^c + z_i]$, with $mc_i \geq 0$. If mc_i is private information to the dealer, this specification is essentially equivalent to assuming that dealers face capital constraint (4) if we are only interested in the size of cost mc_i, or, equivalently, the size of the shadow cost of the capital constraint, $\lambda_i \kappa$.

2.2 Equilibria

We focus on symmetric equilibria since dealers are ex-ante identical.

Definition 1. *A symmetric equilibrium is a collection of demand functions $p^*(\cdot, \theta_i)$ that for each dealer, and almost every θ_i, maximizes expected surplus (2) subject to capital constraint (4). An equilibrium is linear if $\frac{\partial p^*(a,\theta_i)}{\partial a}$ is constant at all a.*

To characterize equilibrium function $p^*(\cdot, \theta_i)$, take the perspective of dealer i with information θ_i, and assume that all other bidders $j \neq i$ play the symmetric equilibrium strategy, $p^*(\cdot, \theta_j)$ or it's inverse, $a^*(\cdot, \theta_j)$. To determine her best-response, $p_i(\cdot, \theta_i)$, dealer i seeks to maximize her expected surplus (2). Following the related literature (e.g., Malamud and Rostek (2017)), we can simplify the maximization problem by leveraging the fact that the asset's return is Normally distributed, $\boldsymbol{R} \sim N(\mu, \sigma)$:

$$\max_{p_i(\cdot,\theta_i)\in\mathcal{B}} \mathbb{E}\left[V_i(\boldsymbol{a_i^c},\theta_i) - p_i(\boldsymbol{a_i^c},\theta_i)\boldsymbol{a_i^c}|\theta_i\right] \text{ with } V_i(\boldsymbol{a_i^c},\theta_i) = \mu[\boldsymbol{a_i^c}+z_i] - \frac{\sigma\rho}{2}[\boldsymbol{a_i^c}+z_i]^2 \text{ subject to}$$

$$\text{capital constraint: } \kappa\mathbb{E}[p_i(\boldsymbol{a_i^c},\theta_i)[\boldsymbol{a_i^c}+z_i]|\theta_i] \leq E_i, \text{ and}$$

$$\text{market clearing: } \boldsymbol{a_i^c} = \boldsymbol{A} - \sum_{j\neq i} a^*(p_i(\boldsymbol{a_i^c},\theta_i),\boldsymbol{\theta_j}), \tag{5}$$

as well as natural boundary conditions.[6] A solution to this problem characterizes a dealer's best-response, which must coincide with the strategy chosen by some other dealer j with the same information as i in a symmetric equilibrium (see Appendix E).

Proposition 1. *In any symmetric equilibrium, dealer i submits demand function, $p^*(\cdot,\theta_i)$, such that $p^*(a,\theta_i) = p$ for all a, given by*

$$p = \frac{v_i(a)}{1+\lambda_i\kappa} - shading(a,p|\theta_i), \tag{6}$$

$$\text{where } v_i(a) = \mu - \sigma\rho[a+z_i] \tag{7}$$

is the dealer's marginal utility from amount a, $\lambda_i \geq 0$ is the Lagrange multiplier of the capital constraint, $\lambda_i[E_i - \kappa\mathbb{E}[p^*(\boldsymbol{a_i^*},\theta_i)[\boldsymbol{a_i^*}+z_i]|\theta_i]] = 0$, and $shading(a,p|\theta_i) = -a(\frac{\partial G(a,p|\theta_i)}{\partial a}/\frac{\partial G(a,p|\theta_i)}{\partial p}) \geq 0$, where $G(a,p|\theta_i) = \Pr(\boldsymbol{A} - \sum_{j\neq i} a^*(p^*(a,\theta_i),\boldsymbol{\theta_j})) \leq a|\theta_i)$ is the probability that dealer i, who bids price $p = p^*(a,\theta_i)$, wins less than a at market clearance given that the other agents play the equilibrium demand $a^*(\cdot,\theta_j)$.

Propositions 3 in the Appendix outlines the equilibrium for discriminatory price auctions

[6] To see why this is the case, re-write (2) as $U(p_i(\cdot,\theta_i)) = \mathbb{E}\left[\mathbb{E}\left[1-\exp(-\rho\omega_i(\boldsymbol{a_i^c},p_i(\boldsymbol{a_i^c},\theta_i))\right]|\theta_i\right]$, where the first expectation is w.r.t. $\boldsymbol{a_i^c}$ and the second expectation is w.r.t. $\boldsymbol{R} \sim N(\mu,\sigma)$. Now insert $\omega_i(\boldsymbol{a_i^c},p_i(\boldsymbol{a_i^c},\theta_i))$ given by (3) and take the expectation w.r.t. $\boldsymbol{R}$ to obtain: $U(p_i(\cdot,\theta_i)) = \mathbb{E}\left[1-\exp(-\rho\{V_i(\boldsymbol{a_i^c},\theta_i) - p_i(\boldsymbol{a_i^c},\theta_i)\boldsymbol{a_i^c}\})|\theta_i\right]$ with $V_i(\boldsymbol{a_i^c},\theta_i) = \mu[\boldsymbol{a_i^c}+z_i] - \frac{\sigma\rho}{2}[\boldsymbol{a_i^c}+z_i]^2$. Given that $1-\exp(-\rho y)$ is strictly increasing for any $y \in \mathbb{R}$, maximizing $U(p_i(\cdot,\theta_i))$ is equivalent to maximizing $\mathbb{E}\left[V_i(\boldsymbol{a_i^c},\theta_i) - p_i(\boldsymbol{a_i^c},\theta_i)\boldsymbol{a_i^c}|\theta_i\right]$. This transformation is valid when adding capital constraint (4).

and step-functions, with the only difference being the shading factor in these alternative settings. In all scenarios, a dealer's true willingness to pay, denoted as $v_i(a)$, is determined by the marginal utility derived from the asset a, given the dealer's gross utility, $V_i(a, \theta_i)$. The dealer's willingness to pay decreases as the amount of the asset increases. For the initial unit of the asset, the dealer obtains the per-unit return μ, while the marginal benefit diminishes for subsequent units based on factors such as the asset's return variance, σ, and the dealer's risk aversion, ρ.

The key insight is that the dealer bids as if participating in a standard multi-unit auction without capital constraints, where their willingness to pay is adjusted as

$$\tilde{v}_i(a) = v_i(a)(1 + \lambda_i \kappa)^{-1}. \tag{8}$$

We refer $\tilde{v}_i(a)$ as pseudo willingness to pay (or pseudo-value). To minimize payments, the dealer shades their pseudo willingness to pay, unless the market is perfectly competitive. For example, in a uniform price auction with smooth functions (considered as the benchmark in Proposition 1), the extent of shading depends on the distribution of the dealer's equilibrium winnings, denoted as $G(a, p|\theta_i)$.

The capital constraint becomes binding when the Lagrange multiplier, λ_i, is strictly positive, while it is slack otherwise. Notably, λ_i is a function of the dealer's private information, θ_i, since it relies on the dealer's expectations of winning, which, in turn, is influenced by the dealer's bidding behavior that is shaped by θ_i.

Deriving sufficiency conditions under which a symmetric equilibrium exists is challenging even for uniform price auctions with smooth demand functions. This is because the slope in the dealer's pseudo willingness to pay changes randomly in the dealer's private information—as if the dealer had private information about her effective degree of risk aversion, $\rho(1 + \lambda_i \kappa)^{-1}$. With random slopes there is no linear equilibrium when dealers have market power. To see why, assume that all dealers other than dealer i submit a linear demand curve. The necessary equilibrium conditions imply that dealer i's best-response is linear if and only if observing price realization p does not update the dealer's belief about the other dealers' constraints. However, even if we assume that this holds for all dealers, the market clearing price is a function of the Lagrange multipliers of all dealers (the full proof is in Appendix E).

Corollary 1. *(i) There is no linear equilibrium when dealers have private information and face capital constraints, unless the market is perfectly competitive. (ii) Under perfect competition, i.e., when $N \to \infty$, dealer i submits her pseudo willingness to pay, $\tilde{v}_i(a)$.*

Corollary 1 implies that the common tools of the related literature—which almost exclusively focuses on linear equilibria to obtain tractability—do not apply. For instance, it is not possible to solve for a dealer's best-response by point-wise maximization. We can, however, solve for a unique symmetric linear equilibrium when abstracting from private information. Alternatively, we could solve for an equilibrium when dealers are asymmetric, for instance, due to different inventory positions, but face no uncertainty. This equilibrium is analogous to Proposition 2, but is not unique—a common feature in uniform price auctions (Klemperer and Meyer (1989)).

Proposition 2. *Let all dealers share the same information with inventory position $z \in \mathbb{R}$, and equity capital $E > 0$. There exists a unique symmetric linear equilibrium in which each dealer submits*

$$p^*(a) = \frac{1}{1+\lambda\kappa}\left(\mu - \rho\sigma\left(\frac{N-1}{N-2}\right)(a+z)\right) \tag{9}$$

$$\text{with } \lambda = \begin{cases} 0 & \text{if } \frac{E}{\kappa} \geq B \\ \frac{B}{E} - \frac{1}{\kappa} > 0 & \text{if } \frac{E}{\kappa} < B \end{cases} \quad \text{with } B = \mu\mathbb{E}\left[\frac{\boldsymbol{A}}{N} + z\right] - \sigma\rho\left(\frac{N-1}{N-2}\right)\mathbb{E}\left[\left(\frac{\boldsymbol{A}}{N} + z\right)^2\right].$$

Intuitively, an equilibrium bid $p^*(a)$ for amount a equalizes the marginal utility (LHS) with the marginal payment (RHS) of the following optimality condition:

$$\mu - \sigma\rho[a + z] = (1 + \lambda\kappa)(p^*(a) + [a + z]\Lambda), \tag{10}$$

where $\Lambda = \frac{1}{N-1}\frac{\sigma\rho}{1+\lambda\kappa}$ measures the dealer's price impact. Λ is known as Kyles' lambda and is 0 when the market is perfectly competitive so that dealers are price-takers. The inverse of the price impact is a common measure of liquidity in exchange markets (e.g., Vayanos and Wang (2013); Malamud and Rostek (2017)).

The marginal utility is the analogue to $v_i(a)$ in Proposition 1. The marginal payment has several components and depends on the regulatory shadow cost of the capital constraint ($\lambda\kappa \geq 0$), and the dealer's price impact, Λ. When the constraint is not binding ($\lambda = 0$) and dealers are price-takers ($\Lambda = 0$), the marginal payment is just the price that the dealer has to pay for amount a. When the constraint binds ($\lambda > 0$) and dealers are price-takers ($\Lambda_i = 0$), the marginal payment is the price they have to pay plus a shadow cost that comes from the capital constraint, which is similar to an ad-valorem tax. When dealers face a binding capital constraint ($\lambda > 0$) and have market power ($\Lambda \neq 0$), Λa measures by how much a dealer's choice impacts the effective price. Not only does this depend on their risk-aversion and the number of players in the market, it also depends on the shadow cost of capital. Finally, when $z \neq 0$, there is an extra term, $\lambda\kappa\Lambda z$, which reflects the regulatory cost that comes from the fact that a dealer's existing inventory, z, is evaluated at the market price in the capital constraint.

To illustrate how to solve for an equilibrium in uniform-price auctions in which bidders face outcome-dependent constraints, we sketch the proof of Proposition 2. A reader who is not interested in technical details, may skip ahead to Section 2.3.

Proof of Proposition 2: We guess that there is a symmetric linear equilibrium, $a^G(p) = \alpha - \beta p$ with $\alpha, \beta > 0$, and assume that all dealers other than dealer i play this equilibrium guess. Dealer i takes the behavior of her competitors as given and chooses points on the residual supply curve $RS_i(p) = \boldsymbol{A} - \sum_{j\neq i} a_j^G(p)$, which shifts randomly only in parallel. This implies that, for ever price p on every demand function that the dealer may submit, there is unique (random) point at which the residual supply curve intercepts the quantity axis: $\boldsymbol{Z} = A - (N-1)\alpha$.

Rather than maximizing over demand functions $p(\cdot)$ that map from prices to quantities directly, it is easier to maximize over bidding functions, $b(\cdot)$, that map from realizations of Z to prices, and then derive the uniquely implied demand function. Imposing market clearance, by inserting $RS(b(Z), Z) = Z + (N-1)\beta b(Z)$ into the objective function, the dealer's maximization problem—the analogue of problem (5)—reads as follows:

$$\max_{b(\cdot)\in\mathcal{B}} \mathbb{E}[U(RS(b(\boldsymbol{Z}),\boldsymbol{Z})) - b(\boldsymbol{Z})RS(b(\boldsymbol{Z}),\boldsymbol{Z})]$$
$$\text{subject to: } \kappa\mathbb{E}[b(\boldsymbol{Z})[RS(b(\boldsymbol{Z}),\boldsymbol{Z}) + z]] \leq E. \tag{11}$$

Abbreviating $b(\cdot)$ by b with derivative b', this problem is equivalent to: $\max_{b\in\mathcal{B}} I(b)$ subject to $L(b) \geq 0$ with $I(b) = \int_{\underline{Z}}^{\overline{Z}} F(b,Z)\phi(Z)dZ$, where $F(b,Z) = [\mu - b][RS(b,Z)+z] - \frac{\sigma\rho}{2}[RS(b,Z)+z]^2$, and $L(b) = E - \int_{\underline{Z}}^{\overline{Z}} H(b,Z)\phi(Z)dZ$, where $H(b,Z) = \kappa b[RS(b,Z)+z]$. Here $\phi(Z)$ is the density function of $\boldsymbol{Z}$ which has support $[\underline{Z}, \overline{Z}]$.

With this, function b^* is optimal if $L(b^*) \geq 0, \lambda L(b^*) = 0, \lambda \geq 0$, $\frac{\partial(F+\lambda H)}{\partial b} - \frac{d}{dZ}\left(\frac{\partial(F+\lambda H)}{\partial b'}\right)$ evaluated at the optimum is 0 for all Z:

$$\mu - \sigma\rho[RS(b^*,Z)+z] = (1+\lambda\kappa)\Big[b^* + [RS(b^*,Z)+z]\Big(\frac{\partial RS(b^*,Z)}{\partial b}\Big)^{-1}\Big], \tag{12}$$

and the natural boundary conditions are satisfied—which is always the case. Note that this condition is equivalent to condition (10). Importantly, density function $\phi(Z)$ does not depend on b. Therefore we can show that function b^* that fulfills the necessary conditions is indeed optimal.[7]

[7]This follows from the fact that $F(b,Z)$ and $K(b,Z) = F(b,Z) + \lambda H(b,Z)$ are for any Z,

From here it is straightforward to solve for an equilibrium, and show that it is unique within the class of symmetric linear equilibria, by matching coefficients of the dealer's best reply in (12) with the equilibrium guess, and showing that these coefficients are unique.[8] In this equilibrium, each dealer wins $\frac{A}{N}$ and the market clears at $P^* = \frac{1}{1+\lambda\kappa}\left(\mu - \sigma\rho\left(\frac{N-1}{N-2}\right)\left(\frac{A}{N}+z\right)\right)$. Depending on the exogenous parameters of the model, the capital constraint either binds or not. When $\frac{E}{\kappa} \geq \mu\mathbb{E}\left[\frac{\boldsymbol{A}}{N}+z\right] - \sigma\rho\left(\frac{N-1}{N-2}\right)\mathbb{E}\left[(\frac{\boldsymbol{A}}{N}+z)^2\right]$, the constraint is not binding, and $\lambda = 0$. Otherwise, the constraint binds and pins down $\lambda > 0$. □

2.3 How capital constraints affect prices and markups

The main prediction of the model is about what happens when the capital constraint is relaxed, for instance, because the minimal capital thresholds decrease. We examine three effects: the impact on price, price impact (and consequently market liquidity), and markup. While all of these effects are derived from our own model and offer novel insights, we particularly emphasize the effects on price impact and markup. These effects are absent in the existing literature, which assumes perfectly competitive markets.

Corollary 2. *Let $P^*(0)$ denote the equilibrium market price when dealers are price-takers and $P^*(\Lambda)$ when they have market power, and consider a relaxation of capital constraints which decreases the shadow costs of capital for all dealers.*

(i) When dealers face only aggregate uncertainty, demand $p^*(\cdot)$ of each dealer i becomes steeper, and market price $P^*(\Lambda)$ increases. Further, the price impact $\Lambda = \frac{1}{N-1}\frac{\sigma\rho}{1+\lambda\kappa}$ of each dealer i, and the markup $= P^*(0) - P^*(\Lambda)$ increase, while market liquidity, $1/\Lambda$, decreases.

and $\lambda \geq 0$, strictly concave as functions of b. Strict concavity implies that $K(b,Z) - K(b^*,Z) < \frac{\partial K(b,Z)}{\partial b}(b-b^*) \leq 0$ for any b and any Z. Multiplying both sides with $\phi(Z)$ and integrating, we see that $\int_{\mathbb{R}} K(b,Z)\phi(Z)dZ < \int_{\mathbb{R}} K(b^*,Z)\phi(Z)dZ$, and similarly for $F(b,Z)$.

[8] For this, note that condition (12) implies that the following equation $\mu - \sigma\rho[a+z] = (1+\lambda\kappa)[p + [a+z]((N-1)\beta)^{-1}]$, which characterizes the dealer's best-reply demand function in the price-quantity space must hold at all a; and that in any symmetric equilibrium the slope of the dealer's best reply must equal to β.

(ii) When dealers have private information, a dealer's demand $\tilde{v}_i(\cdot)$ becomes steeper, and the market price $P^*(0)$ increases if the market is perfectly competitive. When at least some dealers shade their bids due to market power, the effects depend on the distribution of signals and supply, and the number of competing dealers.

Figure 1 illustrates two types of effects from relaxing capital constraints. The first type is non-strategic. Since the effective price, $(1 + \lambda\kappa)p$, decreases, it becomes cheaper for the dealer to buy larger amounts. The dealer's pseudo willingness to pay shifts upward and becomes steeper.[9] As a result, the market price would increase if the market was perfectly competitive, unless supply adjusts. This prediction is in line with He and Krishnamurthy (2012, 2013) and Brunnermeier and Sannikov (2014). In their models, a positive shock to the net worth, i.e., equity capital, of a dealer increases its risk-bearing capacity, which leads to higher asset prices. In our model, risk aversion is constant.

The second type of effect is strategic, and is absent of existing models that feature perfect competition. Dealers have higher price impact, and thus enjoy more market power, when their constraints are relaxed. Higher market power leads to lower market liquidity, and stronger bid-shading, resulting in a larger markup. This pushes down the market price relative to the price that would arise in a perfectly competitive market. When dealers face only aggregate uncertainty, the non-strategic price effect which pushes the price upward dominates the strategic price effect, so that the market clearing price increases when constraints are relaxed.

To understand these effects, it helps to go through how the dealer determines her best response in a simplified environment with complete information. Then, she trades against a known residual supply curve, $RS_i(p) = A - \sum_{j \neq i} a_j(p)$, and chooses the point on that curve that maximizes her own surplus. If the other dealers submit flatter demand curves,

[9] The willingness to pay becomes steeper rather than just shifting in parallel because the capital constraint depends on the nominal, not the real, value of the amount the dealer wins at market clearance. If the constraint was in real values, the willingness to pay would only shift in parallel.

the residual supply curve is flatter. A flatter curve, in turn, implies that the dealer impacts the market clearing price more strongly by her own demand—moving along a flat residual supply curve changes the price more strongly than moving along a steep residual supply curve. The dealer's price impact and the markup increase.

When dealers face uncertainty, a similar logic applies. The difference is that the dealer now trades against a random residual supply. With aggregate uncertainty about the supply, the residual supply shifts randomly in parallel (for any fixed $\lambda \in \mathbb{R}^+$). A dealer now goes through all possible realizations of supply and chooses the optimal demand point on each of them. Each of these realizations is flatter, and hence the dealer's price impact is larger.[10] When there is private information, the residual supply curve moves randomly in arbitrary ways. Therefore, we cannot make a clear prediction on the effect on the price impact and the markup.

Taken together, Corollary 2 highlights that the effect of capital constraints on prices, price impact, and markups depends on whether dealers have access to private information when bidding, and the degree of competition in the auction. When dealers face only aggregate uncertainty or the market is perfectly competitive, we can derive by how much the price, the price impact, and the markup change in response to a change in the shadow cost of capital, i.e., $\lambda\kappa$.

Corollary 3. *(i) When dealers have market power but no private information, a 1% decrease in the shadow cost of capital ($\lambda\kappa$), leads to an increase in the market price, the price impact, and the markup equal to $\eta = \left|\frac{1}{1+\lambda\kappa} - 1\right|\%$. (ii) When the market is perfectly competitive so that the markup and price impact are zero, the market price increases by η when dealers don't have private information, and by $\eta^\infty = \left|\frac{1}{1+\mathbb{E}[\boldsymbol{\lambda_i}]\kappa} - 1\right|\%$ when they do.*

Summarizing, our model helps explain how capital constraints affect asset prices, price impact (and with that liquidity), and markups. When dealers have no private information,

[10]In equilibrium the λ must be consistent with the submitted demand curves.

Figure 1: Non-strategic and strategic effect when capital constraints are relaxed.

(a) Dealer's pseudo willingness to pay $\tilde{v}_i(a)$ (b) Residual Supply$_i(p)$

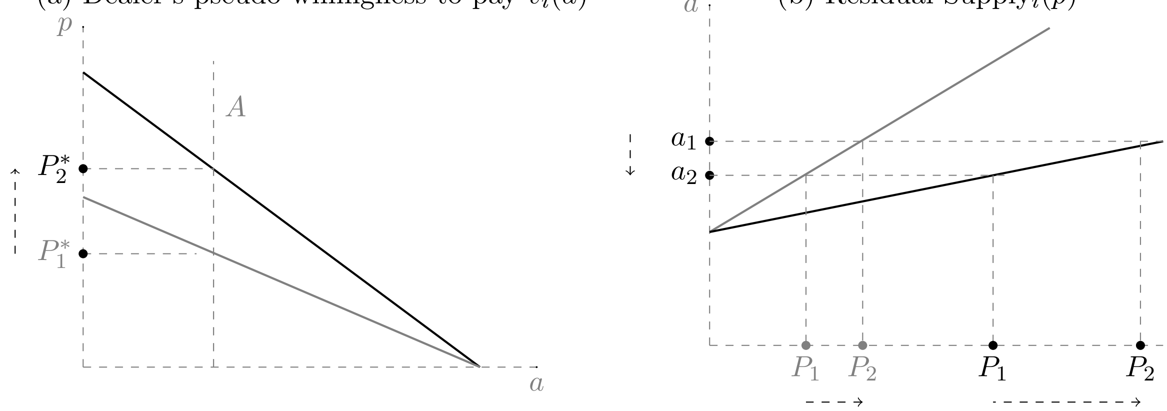


Figure 1 illustrates the change in the dealer's pseudo willingness to pay and her residual supply curve, conditional on one realization of supply, when capital constraints are relaxed in (a) and (b), respectively, for the case without private uncertainty and zero-inventories ($z = 0$). In gray we see the initial pseudo willingness to pay, $\tilde{v}_i(\cdot)$, and residual supply curve, $RS_i(\cdot)$. Demand becomes steeper and the residual supply curve becomes flatter, as shown by the black line, when constraints are relaxed. In (a) we see how this increases the market clearing price, P^*, when supply is fixed in a perfectly competitive market. In (b) we see the increase in the price impact, which measures by how much the clearing price changes, $P_2 - P_1$, when the dealer marginally changes her demand from a_1 to a_2.

their demand becomes steeper, the price, price impact, and markup increases when capital constraints are relaxed. For primary markets, this highlights that relaxing capital constraints increases auction revenues at an implicit cost of larger price distortion. In the context of exchange markets, where higher markups indicate reduced market liquidity, our model reveals a negative side-effect of relaxing capital constraints.

In the presence of private information, it becomes an empirical question whether, and to what extent, capital constraints affect demand, prices, price impact, and markups. In Sections 3 and 4 we illustrate how to use our framework to empirically analyze how prices and markups change when capital constraints change, using data on Canadian Treasury auctions. These auctions utilize the discriminatory price format, where price impact is not well defined. Hence, we exclude price impact (and market liquidity) from the remainder of the paper.

3 Institutional setting and data

Canadian Treasury auctions have the attractive feature that dealers submit entire demand curves, which we can link to balance sheet information of each dealer—a unique feature of our data.

Market players. There are eight deposit-taking primary dealers in Canada who are federally regulated.[11] They dominate the Canadian Treasury market and intermediate the vast majority of the daily trade volume in government bonds. More broadly, these banks dominate the Canadian banking sector and hold over 90% of the sector's assets.

Primary dealers have a responsibility, as market-makers, to buy bonds from the government and trade them with investors, brokers, or one another to provide liquidity. They hold a substantial amount of bonds on their own balance sheets (see Appendix Figure A1). In exchange, primary dealers enjoy benefits, including privileged access to liquidity facilities and overnight repurchase operations at the central bank.

Treasury auctions. Governments issue bonds of different maturities in the primary market via regularly held uniform price or discriminatory price auctions. In Canada, auctions are discriminatory price. Each bidder submits a step-function with at most $K = 7$ steps, which specifies how much a bidder offers to pay for specific amounts of the good for sale. Auctions take place several days a week. Anyone may participate, but most of the supply is purchased by dealers. The largest eight dealers purchase the majority of the Treasury supply in order to sell (or lend) on the secondary market.

[11]In total there are eleven primary dealers. One of these dealers is provincially regulated and two are private securities dealers. They face different capital regulation than the eight dealers we study. We therefore do not observe any balance sheet information for these players. Technically, two of the eight banks have multiple dealers. For example, the Bank of Montreal has two dealers (Bank of Montreal and BMO Nesbitt Burns) who attend different Treasury auctions, and therefore do not compete or share information within an auction. We treat them as one dealer.

Capital constraints. According to a survey among market participants, the Basel III LR represents the most relevant capital constraint when trading government bonds (CGFS (2016)). This regulatory requirement came into effect in September 2014 to reduce systematic risk—a benefit which we do not consider in this paper. We focus on the cost-side of the constraint, which was emphasized by Duffie (2018), He et al. (2022), and others.

Formally, the LR measures a bank's Tier 1 capital relative to its total leverage exposure, and must be at least 3%:

$$\text{LR}_{iq} = \frac{\text{regulatory capital of bank } i \text{ in quarter } q}{\text{total leverage exposure of } i \text{ in } q}.$$

Tier 1 capital consists primarily of common stock and disclosed reserves (or retained earnings), but may also include non-redeemable non-cumulative preferred stock; the leverage exposure includes the total notional of all cash and repo transactions of all securities, including government bonds, regardless of which securities are used as collateral (for more details see OSFI (2023)).

In reality, banks refrain from getting close to the minimal Basel III threshold (see Figure 2b, explained below).[12] One reason for this is that each institution faces an additional supervisory LR threshold that reflects the underlying risk of the bank's operations. Another reason is that banks tend to hold sufficient conservation buffer for Tier 1 capital so as to avoid punishment in the form of restricted distributions (including dividends and share buybacks, discretionary payments and bonus payments to staff).

Regulatory change. To separately identify shadow costs of capital and risk aversion, we rely on a regulatory change that temporarily eliminated the capital constraint for Treasuries. When dealers failed to absorb the extraordinary supply of government bonds in March 2020, government bonds, central bank reserves, and sovereign-issued securities that qualify as High

[12]Barth et al. (2005), Berger et al. (2008) and Brewer et al. (2008) document that bank capital is substantially above the regulatory minimum in countries other than Canada.

Figure 2: The effect of the exemption on Treasury positions and the LR

(a) Aggregated positions in Treasuries (b) Time series of LR for an average bank

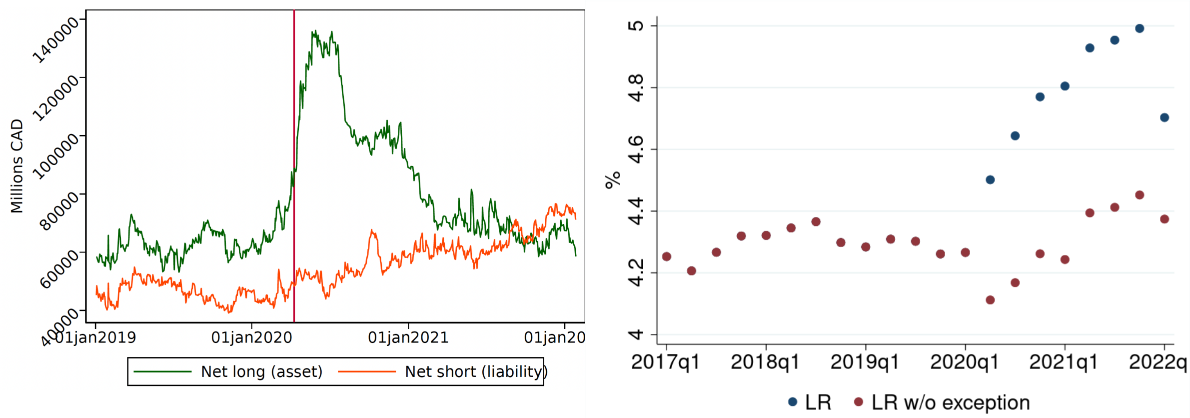


Figure 2a shows the aggregated amount of Canadian government bonds that the biggest six Canadian banks hold in long (in green) and short (in red) positions in millions of C$ from January 2019 until February 2022. The vertical line is April 9, 2020, when government bonds were exempt from LR. Figure 2b shows the time series of the LR (in %) of an average bank. In blue, we show the actual LR. In red the counterfactual LR that the average bank would have had if government bonds, central bank reserves, sovereign-issued securities that qualify as HQLA, and exposures related to the PPP were not exempt. In 2022q1, the LR does not get back to its original level, partially because central bank reserves are still exempted.

Quality Liquid Assets (HQLA) were temporarily exempted from the LR constraint—starting on April 9, 2020.[13] As a result, the LR spiked upward, moving away from the constraint (see Figure 2b). The exemption of government bonds and HQLA ended on December 31, 2021, while reserves continued to be excluded.

Data. We combine multiple data sources. First, we obtain bidding data of all regular government bond auctions between January 1, 2019 and February 1, 2022 from the Bank of Canada. We see how much is issued of which security, and the maturity category of which

[13]Exposures related to the US Government Payment Protection Program (PPP), which are minor in the case of Canadian banks, were also temporarily exempted. The announcement to start the exemption period is available at: `www.osfi-bsif.gc.ca/Eng/fi-if/in-ai/Pages/20200409-dti-let.aspx`; the one to end it is here: `www.osfi-bsif.gc.ca/Eng/fi-if/in-ai/Pages/lrfbunwd.aspx`, both accessed on 05/31/2022.

there are five (2Y, 3Y, 5Y, 10Y and 30Y). We also observe who bids (identified by a legal entity identifier) and all winning and losing bids at auction closure. For consistency, we restrict attention to bids of the eight dealers who are deposit-taking throughout most of the paper.

Second, we collect balance sheet information for these eight dealers at the company holding level. Specifically, we obtain the LRs of each daeler, which is reported quarterly, at the end of January (first quarter), April (second quarter), July (third quarter) and October (fourth quarter of the reporting year) from 2015q1 until 2022q1 from a data source, called LR.[14] In addition, we obtain the daily aggregated long and short positions in government bonds of the six largest dealers from the Collateral and Pledging Report (H4). Finally, we collect information on who holds government bonds—banks versus other investor types—from the National Accounts (Statistics Canada).

Third, we gather information on the volatility of the return, i.e., the price, that a dealer expects to obtain from selling government bonds in the secondary market. For this we leverage the fact that dealers start selling bonds (that are about to be issued at auction) when the tender call opens, which happens one week before the auction closes. They observe the distribution of prices at which they can sell a particular bond before the auction closes, which gives them a precise idea about the return volatility from buying a particular bond at auction and selling it in the secondary market. To also observe this price distribution, we obtain prices (and yields) of essentially all trades with Canadian government bonds from January 1, 2019 until February 1, 2022. These data are collected by the Industry Regulatory Organization of Canada in the Debt Securities Transaction Reporting System (MTRS2.0) and are made available for research with a time lag.

Fourth, we collect the Implied Volatility Index for Canadian Treasuries over the same

[14]One of the banks, HSBC, has a different reporting schedule than the others. Its fiscal year ends in December, instead of October. In our empirical analysis this difference is absorbed when we include dealer fixed effects.

Table 1: Summary statistics

	Mean	Median	Std	Min	Max
Supply (in bn C$)	4.12	4.00	1.23	1.40	7.00
Average bid yield (in %)	1.04	1.09	0.58	0.20	2.18
Years to maturity	8.40	5.03	9.62	2.0	32.62
Number of (deposit-taking) dealers	8	8	0	8	8
Number of steps in demand curve	4.80	5	1.43	1	7
Maximal amount demanded (in % of supply)	7.28	6.25	3.84	0.07	35
Amount dealer won (in % of supply)	6.55	4.98	6.22	0	44.22
Quarterly LR (in %)	4.41	4.36	0.28	-	-
Return volatility (normalized)	1	0.76	0.90	0	7.93

Table 1 shows the average, median, standard deviation, minimum and maximum of key variables in our sample. Our auction data goes from January 1, 2019 until February 1, 2022 and counts 176 bond auctions. The LR data goes from 2015q1 until 2022q1. The min and max LR are empty because we cannot disclose this information.

time period. This index measures the expected volatility of the market over the next 30 days and is based on option prices on short-term interest rate futures (Chang and Feunou (2014)). It is similar to the VIX, which measures volatility in equity markets.

Summary statistics. An overview of the main variables is presented in Table 1.

Note that for our empirical findings we express bond values in yields-to-maturity, rather than prices. This makes the value of bonds that have different maturities and coupon payments more comparable, and implies that demand schedules are increasing.

In line with this convention, we compute the auction-specific return volatilities as standard deviation of yields (expressed in %) at which a dealer sells a bond that is to be auctioned during the week preceeding the auction. To avoid our estimates being driven by the absolute magnitude of the volatility, we normalize the return volatility by its average. Figure 3 shows that the resulting return volatility is similar, yet not identical, to the Implied Volatility Index

Figure 3: Return volatility

(a) Across maturities (b) Over time

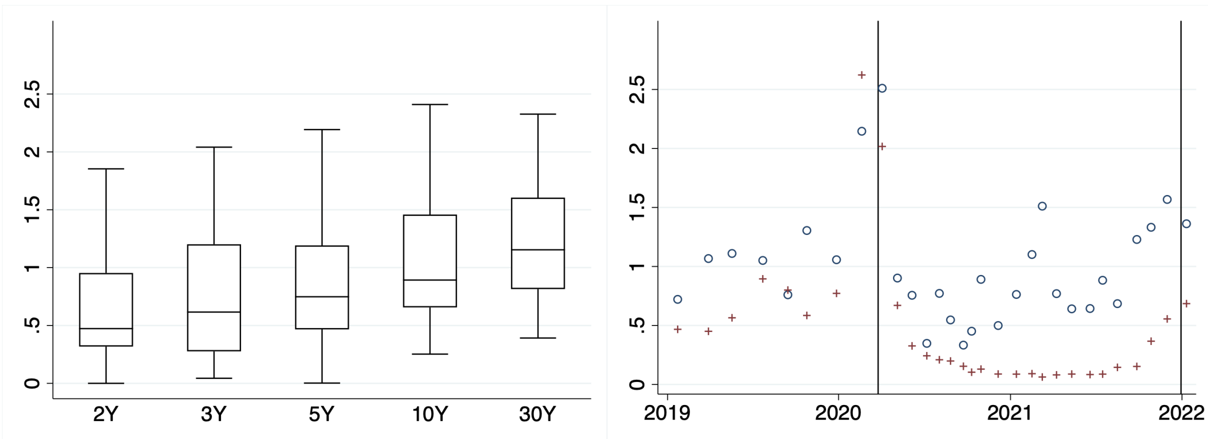


Figure 3a shows the distribution of the normalized return volatility for each maturity category, excluding outliers. Figure 3b shows a binned scatter plot of the return volatility (in circles) and the implied volatility index in % (in pluses) across time. The correlation between these two volatility indices in 0.3. The black lines mark the beginning and end of the exemption period (09 April 2020) and (01 January 2022).

for Canadian Treasuries (in %). One reason for the difference is that dealers sell bonds they buy at auction very quickly, if not before the auction occurs, so that their time horizon is shorter than 30 days.

4 Quantification

To quantify by how much the yield and markups change when capital constraints are relaxed, we adjust the benchmark model to better fit the data generating process.

Model adjustments. Consider an auction t that issues a bond of maturity m=$\{2Y$, $3Y$, $5Y$, $10Y\}$. In line with the institutional setting, the auction is discriminatory price and bidders submit step functions, i.e., sets of $K \geq 1$ quantity-price tuples, $\{a_k, p_k\}_{k=1}^{K}$. Thus, a dealer's equilibrium demand satisfies the condition of Proposition 3 (ii) in the Appendix.

As predicted by our theory, a dealer who draws *iid* private information θ_{ti} from some auction-specific distribution on the day of auction t is willing to offer

$$\tilde{v}_{tik} = f_t(\theta_{ti}) - \beta_{ti}\sigma_t a_{tik}, \text{ with } \beta_{ti} = \frac{\rho_m}{1 + \lambda\kappa_{ti}} \tag{13}$$

for amount a_{tik}, where $f_t(\cdot)$ is some continuous function, for instance, $\mu_t - \rho_m \sigma_t z_{ti}$. Parameter $\rho_m \geq 0$ measures the degree of risk aversion for a bond with maturity m, i.e., we allow, but do not impose, risk aversion to vary in the bond's maturity to reflect the fact that longer bonds may be risker to hold than shorter bonds. Further, parameters $\lambda\kappa_{ti} \geq 0$ represent the shadow costs of capital. Note that we estimate the product of the Lagrange multiplier of the constraint and the capital threshold to avoid having to specify a capital threshold. To highlight this, we relabel the shadow costs $\lambda\kappa_{ti}$ instead of $\lambda_{ti}\kappa$.

Identification and estimation. To identify our main parameters of interest, the shadow costs of the capital constraint, $\lambda\kappa_{ti}$, and the dealer's risk aversion, ρ_m, we proceed in three steps. First, given bids in the auction, we back out how much dealers are truly willing to pay at each step k they submit, i.e., for each submitted amount, under the assumption that all bidders are rational and play the equilibrium.

For this, we need to estimate the distribution of the market clearing price, $\boldsymbol{P_t^*}$, from the perspective of each dealer, θ_{ti}. To do this, we adopt the resampling procedure introduced by Allen et al. (2023), who build on Hortaçsu and Kastl (2012), Kastl (2011), and Hortaçsu and McAdams (2010).

This resampling procedure takes institutional details of Canadian Treasury auctions, which are omitted in our theoretic model, into account, so as to obtain an unbiased estimate of the price distribution. For example, it adjusts for the fact that there are not only dealers, but also customers who bid via dealers. Importantly, these details only affect the way we estimate the price distribution, but not the equilibrium condition itself (for a given price distribution). For example, a dealer's information set, θ_{ti}, includes a customer's bid if the dealer observed a customer's bid before bidding (see Hortaçsu and Kastl (2012) and Allen et al. (2023) for more details).

Once we know all elements of the equilibrium condition, we can solve for the unique pseudo-value, $\tilde{v}_{tik}$, that rationalizes the observed bid in each auction t of each dealer i at each submitted step k (Kastl (2011)).

Second, we fit the model-implied functional form of the dealer's pseudo willingness to pay (13), and estimate the slope coefficients, $\beta_{ti} = \rho_m(1 + \lambda\kappa_{ti})^{-1}$, using variation in the pseudo-values across steps. For this, we use functions with at least two steps, which represent 99% of all functions, and the fact that we observe the volatility term, σ_t.

Finally, we separately identify two sets of parameters with the degree of risk-aversion and shadow costs, by comparing slopes in auctions around the two policy changes, under the assumption that risk-aversion (per maturity class) is constant around the policy changes.[15] Given that capital requirements must be fulfilled quarterly, we use auctions in the 2020q1–2020q2 period when Treasuries were exempt, and auctions in the 2021q4 and 2022q1 period when the exemption period ended. Formally, we find parameters $\{\rho_m, \lambda\kappa_{ti}\}$ for all m, t, i, by fitting

$$\tilde{v}_{tik} = \zeta_{ti} - \sum_{it} \beta_{ti}\mathbb{I}(dealer = i)\mathbb{I}(auction = t)\sigma_t a_{tik} + \epsilon_{tik}, \tag{14}$$

with data from 2020q1–2020q2, and from 2021q4-2022q1, respectively. Here $\beta_{ti} = \rho_m(1 + \lambda\kappa_{ti})^{-1}$ such that $\lambda\kappa_{ti} = 0$ for all dealers i and auctions t when Treasuries are exempt, $\lambda\kappa_{ti} \geq 0$ otherwise, and $\rho_m \geq 0$; ζ_{ti} is a dealer-auction fixed effect, σ_t is the return volatility plotted in Figure (3), and ϵ_{tik} represents (finite sample) measurement error in the pseudo-values.

We express quantities in percentages of auction supply to avoid that changes in the

[15]An alternative would be to set risk-aversion constant across maturities—this is rejected by the data as shown in Figure 6.

supply, which increased substantially during the COVID pandemic, affect our estimates. In Appendix D, however, we also show our estimates when using quantities in absolute terms. With values expressed in percentages of yields, β_{ti} measures by how many percentage points the dealer's pseudo willingness to pay decreases when demand increases by 1% of auctions supply in an auction with average return volatility ($\sigma_t = 1$).

Estimation findings. Before identifying our parameters of interest, we analyze the slope coefficients, β_{ti}, from regression (14), when estimated without constraints, and using all auctions from 2019 onwards as our starting point (see Figure 4). We do this using estimated pseudo-values as well as observed bids.

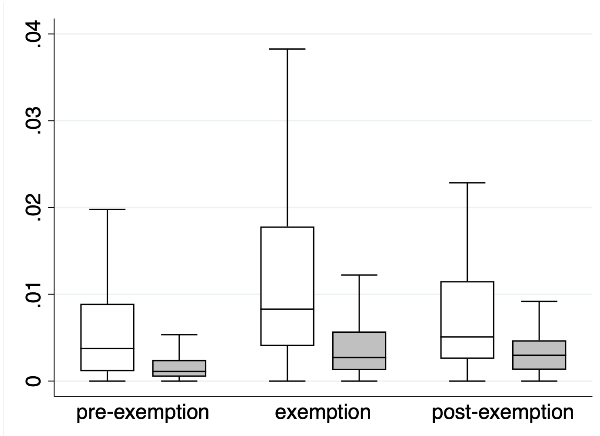
Our theory predicts that true demand curves, i.e., equation (13), becomes steeper when Treasuries are exempt from the constraint if constraints bind. Thus, if the shadow costs of the constraint are non-zero, slope coefficients, β_{ti}, should be larger during the exemption period than in regular times. This would also be true for submitted demand functions, i.e., bids, when competition is sufficiently strong or dealers face little private information.

Indeed, we find that the slope coefficients of both values and bids are higher when Treasuries are exempt. The median slope during the exemption period when using values is lower than when using bids, because shading decreases in quantity. This is shown in Figure 5a and implies that the dealer's pseudo willingness to pay is steeper than the bidding function she submits.

Next, we separate the degree of risk aversion form the shadow costs by estimating regression (14) with constraints using data of auctions around the policy changes.

We find that risk aversion is relatively low for all bond-types with no clear pattern with respect to maturity length (see Figure 6). The median degree of risk aversion is 0.006. This implies that a typical dealer is willing to pay 0.6 basis points less for 1% more of the auction supply in an auction with average return volatility. If dealers were risk neutral, their willingness to pay would be perfectly flat.

Figure 4: Slope coefficients of estimated value and observed bids



The white boxplots of Figure 4 show the distribution of the estimated slopes coefficient of the dealers' pseudo willingness to pay in auction t of regression (14) without imposing restrictions on ρ or $\lambda\kappa_{ti}$ for three time periods: before the exemption of Treasuries from the LR (2019q1–2020q1), during the exemption period (2020q1–2021q4) and after the exemption (2022q1). The gray boxplots show the analogue when using bids instead of estimated values. Dealer values and bids are in %, quantities are in % of auction supply.

In comparison, the existing auction literature estimates risk aversion of similar, yet typically larger, magnitudes in non-financial settings (but given CARA preferences).[16] Most papers consider single-unit auctions. For instance, Bolotnyy and Vasserman (2023) estimate a median degree of risk aversion of firms in procurement auctions to be 0.08. One exception is Häfner (2023), who analyzes discriminatory price auctions for Swiss tariff-rate quotas. He finds that the majority of bidders exhibit a risk aversion parameter of 0.007.

[16]We are not aware of papers that estimate CARA risk aversion for financial intermediaries. The related macro-finance literature on intermediary asset pricing calibrates dynamic models with CARA utility. For instance, He and Krishnamurthy (2013) assume that financial intermediaries have a constant relative risk aversion of 2. More remotely related is a literature that estimates the intertemporal elasticity of substitution, which equals the inverse of CRRA risk aversion (e.g., Vissing-Jørgensen (2002)). However, the CRRA parameters are not comparable to our estimate given that we are relying on CARA preferences.

One way of quantifying the degree of risk aversion is to compute the certainty equivalent. We refrain from doing so, because it is not straightforward in our case. The reason is that there are two layers of uncertainty. The first layer comes from the fact that the asset's return is random. The second layer comes from the fact that the auction outcome is uncertain.

Figure 5: Shading

(a) Per step (b) Average per period

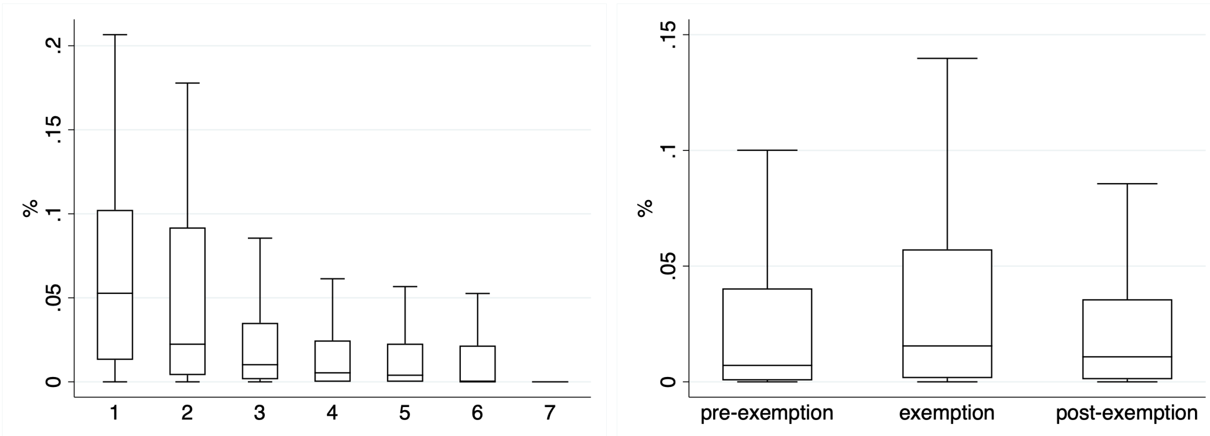


Figure 5a shows box plots of how much dealers shade their bids (in %) at each of the seven steps. It is the difference between the submitted yield bid and the estimated value, both in percentage. The distribution for each step is taken over dealers and auctions. Shading factors are small in absolute terms, and comparable to those in the literature (e.g., Chapman et al. (2007); Kang and Puller (2008); Kastl (2011); Hortaçsu et al. (2018); Allen et al. (2020, 2023)). Figure 5b shows the distribution of shading across auctions, dealers and steps in 2020q1 (pre-exemption), 2020q2 and 2021q4 (exemption), and 2022q1 (post-exemption).

Shadow costs, which are shown in Figure 7, vary substantially across dealers and auctions, reaching higher values in 2020, when dealers struggled to absorb excess supply of Treasuries onto their balance sheets, than in 2022 when markets had calmed down. The long tail in the distribution of shadow costs suggests that there are some auctions in which some dealers expect to take losses. The median shadow cost is 3.5%.

Our cost estimates are in the range of existing estimates found in other markets using different data and different methodologies. For instance, Du et al. (2018) use the overnight spread between the interest rates on excess reserves paid by the Federal Reserve and the Fed Funds as proxy for the shadow costs of bank's balance sheets, which is a couple of basis points. Adrian et al. (2014) fit an augmented Fama-French factor model using quarterly balance sheet data from U.S. security broker-dealers from 1968q1 to 2009q4 and U.S. stock returns. They compute a price of leverage (which proxies for the funding constraint of

Figure 6: Risk aversion bond type

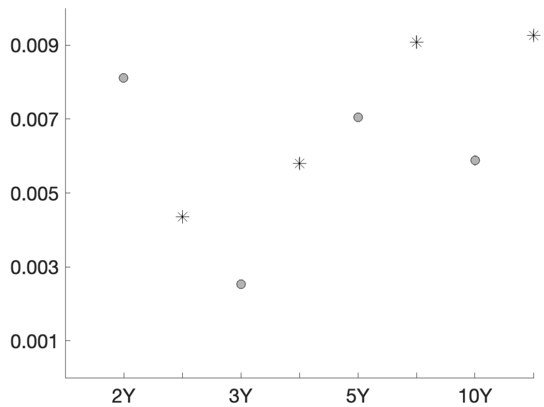


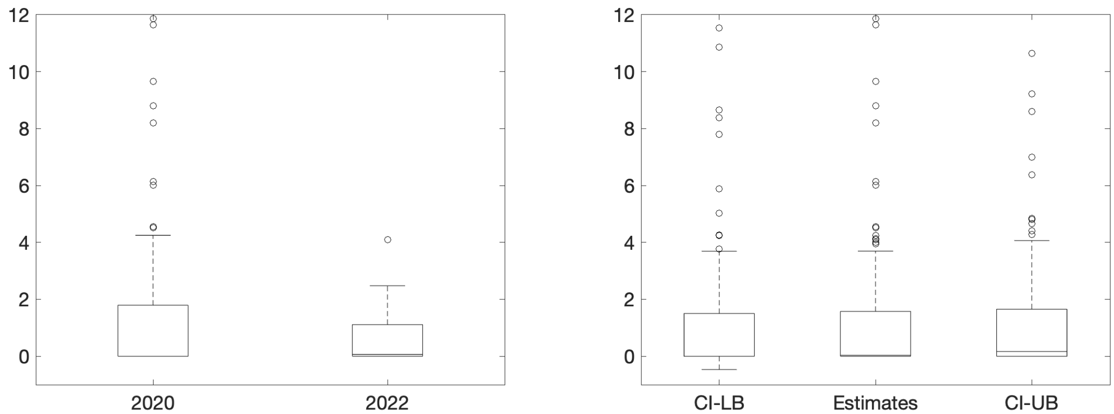
Figure 6 shows the risk aversion estimates, ρ_m, around both policy changes in 2020q1–202q2 (in circles) and in 2021q4–2022q1 (in stars). It also plots the 95% confidence intervals for $m = \{2Y, 3Y, 5Y, 10Y\}$ for 2021q4–2022q1, but given that these intervals are very tight, they are not visible. To compute standard errors and these intervals, we fit equation (14) for each bootstraped estimate of pseudo-values. Each coefficient is measured in % of yield relative to % of supply.

Brunnermeier and Pedersen (2009), among others) of roughly 10% per year.

To get a better sense of how large the median shadow cost is in our setting, we compare it to the typical markup a dealer charges when selling the bond. Specifically, we compute the median difference between the bid yields at auction and the average yield obtained from selling one week before or after the auction. This difference is identical to the median shadow cost, suggesting that dealers barely break even in a typical auction. This is in line with the fact that primary dealers have steadily exited the Canadian Treasury market since the establishment of the primary dealer system, in which primary dealers are responsible for regularly buying Treasuries from the government to actively trade them in the secondary market (see Allen et al. (2023)). This finding also suggests that it might be valuable to carry the primary dealer status to generate revenues outside of the Treasury market.[17]

[17]For instance, primary dealers have access to central bank liquidity facilities and can use their position in the government bond market to cross-sell to investors other investment products, such as underwriting or trading corporate debt. They are also more likely to attract foreign investors

Figure 7: Shadow costs



The LHS of Figure 7 shows a distribution of shadow cost point-estimates, $\lambda\kappa_{ti}$, over auctions t around the two policy changes in 2020 and 2021-2022, excluding outliers. The RHS shows the distribution of the lower and upper bounds of the 95% confidence intervals for each point-estimate (CI-LB and CI-UB, respectively), in addition to the distribution of the point estimates (Estimates) pooled across all auctions, and excluding outliers. The confidence intervals of the shadow costs are bootstraped, analogous to those of the degree of risk aversion.

Discussion. To separate shadow costs from the degree of risk aversion we rely on three main identifying assumptions. Here we discuss what happens when they don't hold.

First, we assume that dealers are rational and play the equilibrium strategy of our empirical auction game. This assumption seems reasonable given that dealers are experienced financial institutions trained to participate in these auctions. Despite this, we also estimate the model under the assumption that dealers bid their true willingness to pay plus a random error term. We find risk aversion is smaller when using bids than when using values because of bid-shading (in line with Figure 4). However, the median shadow cost of 3.1% closely aligns with the median obtained from estimates based on willingness to pay.

Second, we assume that a dealer's pseudo willingness to pay is given by function (13), but this might not be the case. For example, there could be other balance sheet costs faced by banks apart from those arising from the capital constraint. In such a scenario, what we

since the primary dealer status signals trustworthiness and stability.

are actually identifying is the change in the total balance sheet cost as regulations change. To examine this, we could compare the median slope coefficient in Figure 4 across different periods, and would conclude that balance sheet costs were smaller during the exemption period.

Another concern to consider is the potential variation in risk aversion across quarters. This is particularly relevant in 2020, during market turmoil, as opposed to 2021/2022 when markets stabilized and most policies, such as Quantitative Easing, ceased. Therefore, it is reassuring to find shadow cost estimates of similar magnitudes for both periods. However, if we wish to allow for varying degrees of risk aversion across quarters, regression (14) identifies the change in the dealer's "effective" risk aversion, β_{ti}, which may depend on shadow costs and other factors in complicated ways.

Third, we assume that we can observe the volatility of the return that a dealer expects to generate from buying bonds at auction and selling them in the secondary market (from prices at which dealer sell bonds in anticipation of the auction). This assumption seems reasonable as dealers actively trade in anticipation of auctions. However, with finite data there may be measurement error in the observed volatility variable, which could potentially bias our slope estimates downward. If we didn't observe the volatility, we would need to impose more structure on the data, for instance, by taking a stance on the data generating process of secondary market prices. Further, it would make identifying our parameters of interest more challenging. For more details, see Appendix C.

Counterfactual. We could use the model to precisely quantify by how much the auction yield and markups changed because capital requirements were relaxed (tightened). In practice, this involves computing counterfactual bidding step-functions, which is not straightforward. Only recently has Richert (2021, 2022) introduced a numerical method to compute counterfactual bids in multi-unit auctions in which bidders don't face constraints. Even in standard auctions, this method is complex, computationally intense, and requires making assumptions on the distribution of bids, which is endogenous.

As an alternative, we provide a back-of-the-envelope calculation, which leverages our tractable theory of Section 2. In particular, Corollary 3 tells us that the market price (yield) increases (decreases) and the markup increases by $\eta = \left|\frac{1}{1+\lambda\kappa} - 1\right|\%$ when the shadow cost of capital decreases by 1%.[18] The corollary generalizes to discriminatory price auctions with ex-ante identical dealers that face aggregate uncertainty and submit linear demand schedules (excluding the statement on price impact). This does not fit our empirical setting—where dealers submit step functions which are only approximately linear (see Appendix Table A1), and dealers are heterogeneous thanks to private information—perfectly. But we can get a rough sense of magnitudes.

Using the median shadow cost of capital (across all auctions and dealers) the η elasticity is about 0.034% or 3.4 bps. To see what this implies for the auction yield and markup consider the first auction in 2022 after the exemption period ended. This auction cleared at a yield of 1.77%, and the average amount by which a dealer shaded her bid, which approximates the markup due to market power, was roughly 3 basis points. Had the exemption not ended—implying an 100% reduction in the shadow cost of capital—the auction would have cleared at a yield of (1-0.034)1.77% $\approx$ 1.71%, with a markup of (1+0.034)3 bps $\approx$ 3.1bps.

This approximation suggests that the Canadian regulator did not face a quantitatively meaningful trade-off when deciding whether to relax or tighten capital constraints—in addition to the way the LR affects trading in the secondary market and concerns about systematic risk. Relaxing capital constraints decreased yields and increased markups by small amounts. This is in line with the insignificant change in bid shading we observe when the policy changed, as shown in Figure 5b.

[18]Note that this statement hinges on the assumption that volatility is independent of $\lambda\kappa$. In practice, this might not always be the case. For instance, Du et al. (2023a) show that regulatory constraint affects the stochastic discount factor of an intermediary. In this case our back-of-the-envelope calculation neglects the indirect effect that a change in $\lambda\kappa$ has on the price and the markup via a change in the volatility.

Robustness. We conduct a series of robustness checks in Appendix D. For example, we explain what happens when we rely on different volatility indices. We also show that we obtain similar risk-aversion and shadow cost estimates when estimating the model with quantities expressed in absolute terms rather than in percentage of supply, or when including different sets of value functions in the estimation.

5 Implications for intermediary asset pricing

The main focus of this paper is on analyzing the effect of changing capital constraints on the price of an asset and markups that arise due to dealer market power. To draw a closer connection to the intermediary asset pricing literature and inspire future research, we extend our formal analysis in Appendix B to study how intermediary market power affects whether commonly considered intermediary frictions (such as moral hazard or capital constraints) matter for asset prices. Here we only briefly mention the main take aways from this exercise and refer to the appendix for details.

Our first finding highlights that intermediary market power matters for intermediary asset pricing (see Corollary 4). We show that it is not possible to eliminate both moral hazard frictions and frictions that arise due to imperfect competition by hiring a manager who competes for the asset and is paid a fraction of the return that the asset will generate. In this sense, intermediary financing frictions always affect the asset price when the asset market is imperfectly competitive.

Our second finding highlights that capital constraints affect the asset price differently depending on the degree of competition (see Corollary 5). We show that when fewer intermediaries compete for an asset, frictions that arise from capital constraints distort the price more or less strongly, depending on asset-market competition. To be more concrete, consider a market in which many intermediaries compete for the asset. If the number of intermediaries decreases, the asset price moves further away from the price that would arise without capital constraints. The reason is that each intermediary wins more of the asset when fewer of them compete. This increases total exposure, and tightens the capital constraint. The opposite is true in a market with few intermediaries. Now, even though each intermediary wins more, the less competitive auction clears at a sufficiently low price. The price effect dominates the quantity effect and relaxes the constraint.

Taken together, these findings underline that it matters to take imperfect competition into account when analyzing how intermediary frictions affect asset prices, and motivates future research to assess the different degrees of competition across asset markets.

6 Conclusion

This paper studies if and how the capitalization of dealers affects asset prices when dealers have market power. We introduce a model to show that weaker capital requirements lead dealers to demand more of the asset at higher prices but also higher markups. We test the model's prediction and estimate the model with data on Canadian Treasury auctions, where we can link asset demand to balance sheet information of individual intermediaries. Our findings highlight that weaker capital requirements reduce the funding cost of debt but increase market power, and reduce market liquidity.

References

Adrian, T., Etula, E., and Muir, T. (2014). Financial intermediaries and the cross-section of asset returns. *The Journal of Finance*, 69(6):2557–2596.

Adrian, T. and Shin, H. S. (2010). Liquidity and leverage. *The Journal of Financial Intermediation*, 19:418–437.

Adrian, T. and Shin, H. S. (2014). Procyclical leverage and value-at-risk. *The Review of Financial Studies*, 27(2):373–403.

Allen, J., Hortaçsu, A., Richert, E., and Wittwer, M. (2023). Dealer exit and hedge fund entry in the Canadian Treasury market. Mimeo.

Allen, J., Kastl, J., and Wittwer, M. (2020). Interconnected multi-unit auctions: An empirical analysis. Working paper.

Allen, J. and Usher, A. (2020). Investment dealer collateral and leverage procyclicality. *Empirical Economics*, 58(2):489–505.

Allen, J. and Wittwer, M. (2023). Centralizing over-the-counter markets? *Journal of Political Economy*, forthcoming.

An, Y. and Song, Z. (2023). Does the Federal Reserve obtain competitive and appropriate prices in monetary policy implementation? *The Review of Financial Studies*, 00(1-45).

Ang, A., Gorovyy, S., and Inwegen, G. B. (2011). Hedge fund leverage. *Journal of Financial Economics*, 102(1):102–126.

Attanasio, O. P. and Weber, G. (1989). Intertemporal substitution, risk aversion and the Euler equation for consumption. *The Economic Journal*, 99(395):59–73.

Ausubel, L. M., Cramton, P., Pycia, M., Rostek, M., and Weretka, M. (2014). Demand reduction and inefficiency in multi-unit auctions. *The Review of Economic Studies*, 81:1366–1400.

Bajari, P. and Hortaçsu, A. (2005). Are structural estimates of auction models reasonable? Evidence from experimental data. *Journal of Political Economy*, 113(4):703–741.

Baron, M. and Muir, T. (2022). Intermediaries and asset prices: International evidence since 1870. *The Review of Financial Studies*, 35(5):2144–2189.

Barth, J., Caprio, G., and Levine, R. (2005). *Rethinking Bank Regulation: Till Angels Govern*. Cambridge University Press.

Bergemann, D., Heumann, T., and Morris, S. (2021). Information and market power. *RAND Journal of Economics*, pages 125–150.

Berger, A. N., DeYoung, R., Flannery, M. J., Lee, D., and Öztekin, Ö. (2008). How do large banking organizations manage their capital ratios? *Journal of Financial Services Research*, 34:123–149.

Bolotnyy, V. and Vasserman, S. (2023). Scaling auctions as insurance: A case study in infrastructure procurement. *Econometrica*, forthcoming.

Brewer, E., Kaufman, G. G., and Wall, L. D. (2008). Bank capital ratios across countries: Why do they vary? *Journal of Financial Services Research*, 34:177–201.

Brunnermeier, M. K. and Pedersen, L. H. (2009). Market liquidity and funding liquidity. *The Review of Financial Studies*, 22(6):2201–2238.

Brunnermeier, M. K. and Sannikov, Y. (2014). A macroeconomic model with a financial sector. *American Economic Review*, 104(2):379–421.

Campo, S. (2012). Risk aversion and asymmetry in procurement auctions: Identification, estimation and application to construction procurements. *Journal of Econometrics*, 1:96–107.

Campo, S., Guerre, E., Perrigne, I., and Vuong, Q. (2011). Semiparametric estimation of first-price auctions with risk-averse bidders. *Review of Economic Studies*, 78:112–147.

CGFS (2016). CGFS papers no 55: Fixed income market liquidity. *Bank of International Settlements*.

Chang, B. and Feunou, B. (2014). Measuring uncertainty in monetary policy using implied volatility and realized volatility. *Bank of Canada Review*, pages 32–41.

Chapman, J. T. E., Mc-Adams, D., and Paarsch, H. J. (2007). Bounding revenue comparisons across multi-unit auction formats under ε-best response. *American Economic Review*, 97(2):455–458.

Check, H., Joslin, S., and Ni, S. (2019). Demand for crash insurance, intermediary constraints, and risk premia in financial markets. *The Review of Financial Studies*, 32(228-265).

Chen, S., Kaniel, R., and Opp, C. C. (2023). Market power in the securities lending market. Working paper.

Corbae, D. and D'Erasmo, P. (2021). Capital buffers in a quantitative model of banking industry dynamics. *Econometrica*, 86(6):2975–3023.

Du, S. and Zhu, H. (2017). What is the optimal trading frequency in financial markets? *The Review of Economic Studies*, 84(4):1606–1651.

Du, W., Hébert, B., and Hubert, A. W. (2023a). Are intermediary constraints priced? *The Review of Financial Studies*, 36:1464–1507.

Du, W., Hébert, B., and Li, W. (2023b). Intermediary balance sheets and the treasury yield curve. Working paper.

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

Duffie, D. (2018). Financial regulatory reform after the crisis: An assessment. *Management Science*, 64:4835–4857.

Elenev, V., Landvoigt, T., and Nieuwerburgh, S. V. (2021). A macroeconomic model with financially constrained producers and intermediaries. *Econometrica*, 89(3):1361–1418.

Fontaine, J., Garcia, R., and Gungor, S. (2022). Intermediary leverage shocks and funding conditions. Working paper.

Gârleanu, N. and Pedersen, L. H. (2013). Dynamic trading with predictable returns and transaction costs. *Journal of Finance*, 68(6):2309–2340.

Gospodinov, N. and Robotti, C. (2021). Common pricing across asset classes: Empirical evidence revisited. *Journal of Financial Economics*, 140:292–324.

Gromb, D. and Vayanos, D. (2002). Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of Financial Economics*, 66:361–407.

Guerre, E., Perrigne, I., and Vong, Q. (2009). Nonparametric identificaiton of risk aversion in first-price auctions under exclusion restrictions. *Econometrica*, 77(4):1193–1227.

Guerre, E., Perrigne, I., and Vuong, Q. (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68(3):525–574.

Gupta, S. and Lamba, R. (2017). Treasury auctions during a crisis. Working paper.

Haddad, V. and Muir, T. (2021). Do intermediaries matter for aggregate asset prices? *The Journal of Finance*, 76(6):2719–2761.

Häfner, S. (2023). Risk aversion in share auctions: Estimating import rents from TRQs in Switzerland. *Quantitative Economics*, 14:419–470.

He, Z., Kelly, B., and Manela, A. (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics*, 126(1):1–35.

He, Z. and Krishnamurthy, A. (2012). A model of capital and crises. *The Review of Economic Studies*, 79(2):735–777.

He, Z. and Krishnamurthy, A. (2013). Intermediary asset pricing. *American Economic Review*, 103(2):732–70.

He, Z. and Krishnamurthy, A. (2018). Intermediary asset pricing and the financial crisis. *Annual Review of Financial Economics*, 10:173–97.

He, Z., Nagel, S., and Song, Z. (2022). Treasury inconvenience yields during the COVID-19 crisis. *Journal of Financial Economics*, 143(1):57–79.

Holstrom, B. and Tirole, J. (1997). Financial intermediation, lonable funds, and the real sector. *The Quarterly Journal of Economics*, 112(3):663–691.

Hortaçsu, A. and Kastl, J. (2012). Valuing dealers' informational advantage: A study of Canadian Treasury auctions. *Econometrica*, 80(6):2511–2542.

Hortaçsu, A., Kastl, J., and Zhang, A. (2018). Bid shading and bidder surplus in the U.S. Treasury auction. *American Economic Review*, 108(1):147–69.

Hortaçsu, A. and McAdams, D. (2010). Mechanism choice and strategic bidding in divisible good auctions: An empirical analysis of the Turkish Treasury auction market. *Journal of Political Economy*, 118(5):833–865.

Huang, W., Ranaldo, A., Schrimpf, A., and Somogyi, F. (2023). Constrained liquidity provision in currency markets. Working paper.

Huber, A. W. (2023). Market power in wholesale funding: A structural perspective from the triparty repo market. *Journal of Financial Economics*, 149(2):235–259.

Jamilov, R. (2021). A macroeconomic model with heterogeneous banks. Working paper.

Kang, B.-S. and Puller, S. L. (2008). The effect of auction format on efficiency and revenue in divisible goods auctions: A test using Korean treasury auctions. *The Journal of Industrial Economics*, 56(2):290–332.

Kargar, M. (2021). Heterogeneous intermediary asset pricing. *Journal of Financial Economics*, 141(2):505–532.

Kastl, J. (2011). Discrete bids and empirical inference in divisible good auctions. *The Review of Economic Studies*, 78:974–1014.

Kastl, J. (2012). On the properties of equilibria in private value divisible good auctions with constrained bidding. *Journal of Mathematical Economics*, 48:339–352.

Klemperer, P. D. and Meyer, M. A. (1989). Supply function equilibria in oligopoly under uncertainty. *Econometrica*, 57(6):1243–1277.

Kyle, A., Obzhaeva, A. A., and Wang, Y. (2017). Smooth trading with overconfidence and market power. *Review of Economic Studies*, 85(1):611–662.

Kyle, A. S. (1989). Informed speculation with imperfect competition. *The Review of Economic Studies*, 56(3):317–355.

Luo, Y. and Takahashi, H. (2023). Bidding for contracts under uncertain demand: Skewed bidding and risk sharing. *RAND Journal of Economics*, page forthcoming.

Malamud, S. and Rostek, M. (2017). Decentralized exchange. *American Economic Review*, 107(11):3320–3362.

Moreira, A. and Savov, A. (2017). The macroeconomics of shadow banking. *The Journal of Finance*, 72(6):2381–2432.

OSFI (2023). Leverage requirements guideline. `http://www.osfi-bsif.gc.ca/Eng/fi-if/rg-ro/gdn-ort/gl-ld/Pages/LR22.aspx`, accessed on 01/13/2023.

Pedersen, L. H. and Gârleanu, N. (2011). Margin-based asset pricing and deviations from the law of one price. *The Review of Financial Studies*, 24(6):1980–2022.

Richert, E. (2021). Eric richertindirect inference techniques as an alternative to direct computation of equilibrium bid distributions. Working paper.

Richert, E. (2022). Wuantity commitments in multiunit auctions: Evidence from credit event auctions. Working paper.

Rostek, M. and Weretka, M. (2012). Price inferences in small markets. *Econometrica*, 80(2):687–711.

Rostek, M. and Weretka, M. (2015). Dynamic thin markets. *The Review of Financial Studies*, 28(10):2946–2992.

Rostek, M. J. and Yoon, J. H. (2021). Exchange design and efficiency. *Econometrica*, 89(6):2887–2928.

Vayanos, D. (1999). Strategic trading and welfare in a dynamic market. *The Review of Economic Studies*, 66:219–254.

Vayanos, D. and Wang, J. (2013). Market liquidity: Theory and empirical evidence. In: The Handbook of the Economics of Finance, G. M. Constantinides, M. Harris, and R. Stulz (eds.), North-Holland Press, Volume 2b, Chapter 19.

Villa, A. T. (2022). Credit misallocation and macro dynamics with oligopolistic financial intermediaries. Working paper.

Vissing-Jørgensen, A. (2002). Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy*, 110(4):825–853.

Vives, X. (2011). Strategic supply function competition with private information. *Econometrica*, 76(6):1919–1966.

Wallen, J. (2022). Markups to financial intermediation in foreign exchange markets. Working paper.

Wang, Y., Whited, T. M., Whu, Y., and Xiao, K. (2022). Bank market power and monetary policy transmission: Evidence from a structural estimation. *The Journal of Finance*, 77(4):2093–2141.

Wilson, R. (1979). Auctions of shares. *The Quarterly Journal of Economics*, 93(4):675–689.

Wittwer, M. (2018). *Pay-as-bid auctions in theory and practice*. PhD thesis, European University Institute.

Wittwer, M. (2021). Connecting disconnected financial markets? *American Economic Journal: Microeconomics*, 13(1):252–282.

Zhang, A. L. (2022). Competition and manipulation in derivative contract markets. *Journal of Financial Economics*, 144(2):369–413.

ONLINE APPENDIX

Intermediary Market Power and Capital Constraints

by Jason Allen, and Milena Wittwer

Appendix A generalizes our benchmark model to incorporate private information in uniform price and discriminatory price auctions. Appendix B generalizes the model to draw implications for the intermediary asset pricing literature. Appendix C explains why it is useful to observe return volatility. Appendix D presents our robustness analysis. Proofs are in Appendix E.

A Discriminatory price auctions and step-functions

Here we adjust our benchmark model to the case of discriminatory price auctions, in which bidders pay the prices they offered to pay for all units won, rather than the market clearing price. We consider two settings, one in which dealers submit continuous demand functions, as in our benchmark model, and one in which dealers must submit step functions as in Kastl (2012). With slight abuse of notation we use the same notation for continuous demand curves, $a_i(\cdot, \theta_i)$, the probability that dealer i who bids price $p = p(a, \theta_i)$ wins less than a at market clearance given that other dealers play the equilibrium demand $a^*(\cdot, \theta_j)$, $G(a, p|\theta_i) = \Pr(\boldsymbol{A} - \sum_{j \neq i} a^*(p(a, \theta_i), \boldsymbol{\theta_j})) \leq a|\theta_i)$, and the Lagrange multiplier, λ_i, even though all of these are auction-format specific.

Proposition 3.

(i) In any symmetric equilibrium with continuous demand curves, dealer i submits demand functions, $p^*(\cdot, \theta_i)$, such that $p^*(a, \theta_i) = p$ for all a, given by

$$p = \frac{v_i(a)}{1 + \lambda_i \kappa} - shading(a, p|\theta_i), \tag{15}$$

where $v_i(a)$ *and* $\lambda_i \geq 0$ *are as in Proposition 1, and* $shading(a, p|\theta_i) = -\frac{1-G(a,p|\theta_i)}{\frac{\partial G(a,p|\theta_i)}{\partial p}} \geq 0$.

(ii) When demand curves are step functions $\{a_k, p_k\}_{k=1}^{K_i}$*, dealer i's equilibrium function satisfies*

$$p_k = \frac{v_i(a)}{1 + \lambda_i \kappa} - \frac{\Pr(p_{k+1} \geq \boldsymbol{P^*}|\theta_i)}{\Pr(p_k > \boldsymbol{P^*} > p_{k+1}|\theta_i)} \tag{16}$$

at every step but the last one; at the last step, the dealer bids truthfully.

The equilibrium condition is like the condition in a uniform price auction, with a different shading factor, which comes from the fact that dealers pay the bids they place for all units won rather than the market clearing price.

When dealers only face aggregate uncertainty and submit continuous demand functions, we can solve for a symmetric equilibrium. For this we assume that all dealers share the same information with inventory position $z \in \mathbb{R}$, and equity capital $E > 0$. Further, we let supply follow a Generalized Pareto distribution, so that the amount a dealer wins in equilibrium, $\frac{\boldsymbol{A}}{N}$, has CDF: $1 - (\frac{\nu + \xi A/N}{\nu})^{-\frac{1}{\xi}}$, with $\xi < \frac{N-1}{N}, \nu = -\xi(\frac{\bar{A}}{N})$.

Proposition 4. *Let all dealers share the same information with inventory position* $z \in \mathbb{R}$*, and equity capital* $E > 0$*. There exists a symmetric linear equilibrium in which each dealer submits*

$$p^*(a) = \frac{1}{1 + \lambda\kappa}\left(\bar{\mu} - \rho\sigma\left(\frac{N-1}{N(1-\xi)-1}\right)(a+z)\right) \text{ with } \lambda = \begin{cases} 0 & \text{if } \frac{E}{\kappa} \geq B \\ \frac{B}{E} - \frac{1}{\kappa} > 0 & \text{if } \frac{E}{\kappa} < B, \end{cases}$$

where $B = \mathbb{E}\left[\frac{(\mu + N\nu\rho\sigma + \mu N(\xi-1))z}{1+N(\xi-1)} + \rho\sigma z^2 - \boldsymbol{A}\left(\frac{\mu + N\nu\rho\sigma + \mu N(\xi-1)}{N(1+N(\xi-1))} - \frac{\rho\sigma(2+N(\xi-2))z}{N(1+N(\xi-1))}\right) + \boldsymbol{A}^2\frac{(N-1)\rho\sigma}{N^2(1+N(\xi-1))}\right]$,
and $\bar{\mu} = \frac{\mu + \mu N(\xi-1) + \rho\sigma(1+N(N-2-N\xi+\nu))z}{1+N(-1+\xi)}$.

The demand functions are analogues to those in Propositions 1 and 2. Therefore Corollary 2 generalizes to discriminatory price auctions; with the exception of the statements about price impact and market liquidity, which are not clearly defined in discriminatory price auctions.

B Intermediary asset pricing implications

Here we study whether the degree of competition between intermediaries affects the way intermediary frictions—specifically, moral hazard and capital constraints—affect asset prices. Technically, we rely on a simple version of the He and Krishnamurthy (2012, 2013) models, presented in He and Krishnamurthy (2018), that builds on Holstrom and Tirole (1997). Our contribution is to introduce imperfect competition in the asset market by relying on insights from the literature on auctions and market microstructure. We consider the simplest auction environment without private signals and known supply. It is straightforward to introduce supply uncertainty as in Proposition 2.

Model with moral hazard. The economy runs for three periods, $t = 0, 1, 2$.[19] There is one risky asset of aggregate supply A that pays out a return $R \sim N(\mu, \sigma)$ per unit, and a numeraire (cash). The return is unknown to all agents in all periods but the last one.

There are $2 < N < \infty$ intermediaries (banks), indexed by I. Each bank serves a unit mass of households (H) who never consider switching banks, i.e., there are fixed households-bank pairs. In addition, each bank has a trading desk i who is responsible for trading the risky asset.

Households cannot directly invest in the asset market, but must invest via a bank. For this, households and their bank contract with trading desk i (of the bank that serves the households) who invests in the risky asset on the households' behalf.

Banks and households have CARA preferences, that is, holding wealth ω_j generates the following utility for an agent of type $j \in \{H, I\}$:

$$u_j(\omega_j) = 1 - \exp\left(-\rho_j \omega_j\right), \tag{17}$$

with risk aversion $\rho_j > 0$. A trading desk and its bank share the same utility function.

[19] Alternatively, we could merge $t = 0$ and $t = 1$ into a single period.

Wealth comes from buying and holding the asset. For instance, if agent j gets a_j at price p, the wealth is $\omega_j = a_j(R - p)$.

The sequence of events is as follows: In period 0, each households-bank pair chooses what fraction ϕ_i of the total wealth (that will be generated from investing in the asset) will be paid to trading desk i in period 2.[20] The contract is chosen to maximize their joint expected utility obtained at the end of the game. Alternatively, you may think of a market designer who chooses ϕ_i's to maximize expected welfare of the economy subject to incentive constraints. In period 1, all N trading desks compete in a uniform price auction to buy a_i of the risky asset, submitting continuous and strictly decreasing demand functions: $a_i(\cdot)$. Each trading desk may decide to shirk or exert effort; $s_i \in \{0, 1\}$, where s_i=1 is shirking. When a trading desk chooses to shirk, the wealth of its bank falls by Δ, but the trading desk gains a private benefit of b. In period 2, the asset pays its return and supply realizes. All transactions take place.

Proposition 5. *Define $m = \frac{\Delta}{b} - 1 \geq 0$. There exists an equilibrium in which $\phi_i = \phi = \frac{1}{1+m}$, and the clearing price is*

$$P^* = \mu - \left(\frac{\sigma\rho_I}{1+m}\right)\left(\frac{N-1}{N-2}\right)\frac{A}{N}. \tag{18}$$

Trading desk i buys amount $\frac{A}{N}$, its bank obtains $\phi\frac{A}{N}$, and each mass of households receives $(1-\phi)\frac{A}{N}$.

In this equilibrium, the market clearing price has the familiar functional form of a uniform price auction with N bidders (here trading desks). $m = \frac{\Delta}{b} - 1$ is the maximum amount of dollars that households can invest (per dollar that the trading desk purchases) so that the trading desk exerts effort in the auction. If the moral hazard friction is small, which happens

[20] We could let the pair choose an affine contract parametrized by (K_i, ϕ_i), where ϕ_i is the linear share of the return generated by the investment that is paid to the trading desk, and K_i is a management fee that is paid to the trading desk independent of the return. In the case of CARA preferences, the lack of a wealth effect implies that K_i plays no role in asset demand and equilibrium prices.

when the benefit b from shirking is small, the trading desk can be incentivized to exert effort with little skin in the game, that is, with a small ϕ_i. The more beneficial it becomes to shirk, the higher ϕ_i must be.

When the asset market is perfectly competitive, as in He and Krishnamurthy (2018), there are two cases depending on how attractive it is for the trading desk to shirk. In the first case, shirking is attractive so that the constraint that incentivizes trading desk i to exert effort, $\phi_i \Delta \geq b$, binds. As a result, the intermediation frictions affect the asset price. In the second case, the incentive constraint doesn't bind, and the first-best solution can be obtained through choosing the optimal contract ϕ_i.

When the asset market is imperfectly competitive, intermediation frictions always affect the asset price. Intuitively, this is because one instrument (per households-bank pair), ϕ_i, cannot correct two frictions: moral hazard and imperfect competition.

Corollary 4. *There is no contract $\phi_i = \phi \ \forall i$ that implements the price and allocation of a frictionless market, in which both banks and households have access to and compete in an auction that induces truthful bidding, that is, avoids bid shading.*

Model with capital constraints. So far, the intermediation friction came from moral hazard. Now we add capital constraints. Suppose that each trading desk purchases $a_i(p)$ of the asset if the asset market clears at price p, and makes loans L to an un-modeled sector of the economy, which we normalize to 0 w.l.o.g. The desk is subject to a Basel III-type capital constraint: $\kappa p a_i(p) \leq E$, where E denotes the total equity capital.

From He and Krishnamurthy (2018) we know that the capital constraint binds only if the moral hazard incentive constraint binds. Given this, it is not surprising, that the auction clearing price is analogous to the price of Proposition 2, where dealers face a similar capital constraint.

Proposition 6. *In equilibrium $\phi_i = \frac{1}{1+m}$ for all i, and the market clears at*

$$P^* = \frac{1}{1+\lambda\kappa}\left(\mu - \left(\frac{\sigma\rho_I}{1+m}\right)\left(\frac{N-1}{N-2}\right)\frac{A}{N}\right),$$

$$\text{with } \lambda = \begin{cases} 0 & \text{if } \frac{E}{\kappa} \geq B \\ \frac{B}{E} - \frac{1}{\kappa} > 0 & \text{if } \frac{E}{\kappa} < B \end{cases} \text{ with } B = \mu\frac{A}{N} - \left(\frac{\sigma\rho_I}{1+m}\right)\left(\frac{N-1}{N-2}\right)\frac{A}{N}^2. \qquad (19)$$

Trading desk i buys amount $\frac{A}{N}$, its bank obtains $\phi\frac{A}{N}$, and each mass of households receives $(1-\phi)\frac{A}{N}$.

Does competition matter? We now analyze whether imperfect competition in the asset market matters for whether and how the asset price is affected by intermediary frictions.

To vary the degree of competition, we vary the number of banks (or trading desks) who compete for the asset. More bidders in an auction translates into greater competition.[21]

Corollary 5. *Define $\bar{N} : \mu = \frac{(4+\bar{N}(2\bar{N}-5))\phi\rho_I\sigma A}{(\bar{N}-2)^2\bar{N}}$ and let $\mu > 0$.*

(i) Intermediary financing frictions always affect the asset price.

(ii) Let the asset market become less competitive in that N decreases to N'. For $N' \geq \bar{N}$, the shadow cost of the capital constraint increases, so that the market price is more strongly affected by the capital constraint. For $N' < \bar{N}$, the shadow cost decreases and the market price is less strongly affected by the capital constraint.

[21]Note that this is different from the main text, in which we measure competition by the extent to which the market price that arises in the market with market power differs to the price that would arise in a perfectly competitive market. Measuring this price wedge directly gives a more precise idea of the impact of market power on prices than counting the number of market participants. However, since this wedge is endogenous, it is less useful for analyzing how changes in market power affect prices. Crucially, both measures of competition are qualitatively identical in that the price wedge (and the price impact) decreases monotonically when the number of market participants increases.

Competition matters in two ways. First, with imperfectly competitive asset markets, it is no longer the case that intermediation frictions—either moral hazard or capital constraints—can be corrected by choosing intermediary remuneration, ϕ_i, optimally. To overcome (or at least reduce) the extra friction which arises from the fact that the asset market isn't perfectly competitive, a more complex remuneration scheme would be necessary.

Second, when the market is less competitive as a result of fewer intermediaries competing for the asset, the shadow cost of the capital constraint changes. Intuitively, a positive shadow cost guarantees that the capital constraint binds: $\kappa P^* a_i^*(P^*) \leq E$. The shadow cost is higher, the larger $P^* a_i^*(P^*)$ would be relative to E in a setting without the constraint. Thus, to understand how the shadow cost changes, we must think through how $P^* a_i^*(P^*)$ changes as the number of bidders N decreases.

There are two opposing effects. On the one hand, each bidder wins more: $a_i^*(P^*) = \frac{A}{N}$ increases. On the other hand, the less competitive auction clears at a lower price: P^* decreases. When the quantity effect dominates the price effect, the shadow cost increases as N decreases. Whether this is true or not depends on how competitive the market is, i.e., the number of bidders. If the degree of competition is sufficiently strong ($N' \geq \bar{N}$), the quantity effect dominates, otherwise ($N' < \bar{N}$) the price effect dominates.

C Return volatility

We illustrate why observing volatility, σ_t, facilitates identification by means of an example.

Recall that we construct return volatility from the prices, p^S_{tij}, a dealer charges when selling a to-be-issued bond to a trader j prior to the auction, that is, before observing her value realization, $\tilde{v}_{ti}(\cdot)$, on auction day: $\sigma_t = Var(p^S_{tij})$. Further, refer to the distribution of $\tilde{v}_{ti}(\cdot)$, specified in equation (13), by F^v_t and the distribution of the shadow cost (which is a random variable given its dependence on θ_{ti}) by $F^{\lambda\kappa}_t$.

Assume that we do not observe σ_t. Then we need some structure on the data generating process to identify our parameters of interest. Here we provide one example of this process, which is by no means exclusive.

Each dealer charges a common price to its clients, that may depend on the value distribution, F_t^v, and a trade-specific markup. This markup depends on the shadow cost of capital that the dealer faces at that moment. To formalize this idea, let dealer i observe a realization of the shadow cost, $\lambda\kappa_{tij}$, from the distribution $F_t^{\lambda\kappa}$ before selling to buyer j. The dealer charges a markup of $\bar{\eta}_t(\lambda\kappa_{tij})$, where $\bar{\eta}_t(\cdot)$ maps the shadow cost draw into $\mathbb{R}$. To identify our parameters of interest, we would need to replace σ_t in equation (13) by $Var(\bar{\eta}_t(\boldsymbol{\lambda\kappa_{tij}}))$. From here we see that identification becomes challenging. For instance, we would need to specify a functional form for $\bar{\eta}_t(\cdot)$, and estimate a system of equations, which includes equation (13), the distribution of shadow costs, and function $\bar{\eta}_t(\cdot)$.

D Robustness analysis

We conduct a series of robustness checks to validate our risk-aversion and shadow cost estimates. All risk-aversion estimates are presented in Appendix Table A2; Appendix Figure A2 shows the distribution of shadow cost estimates for all specifications.[22]

We start by analyzing the sensitivity of our parameter estimates to the number of steps included in the values functions. In our benchmark specification, we include all functions with at least 2 steps (which are essentially all functions) to avoid a potential bias coming from omitting functions. Given that we linearly interpolate between steps using our model, we might be concerned about doing this when there are few steps. Our results, however, are robust to using value functions with more steps—3 to 6, where we do not include robustness for 7 steps since not all dealers use the maximum allowable number of steps in all auctions.

Next, we estimate equation (14) with quantities expressed in million C\$. In our benchmark specification, we normalize quantities by the auction supply to avoid our estimates

[22] A robustness analysis of the value estimates is provided in Allen et al. (2023).

being affected by the fact that the Bank of Canada issued larger amounts of debt during the exemption period than in regular times. Given that dealers have an obligation to actively participate in the auctions, the increased supply implies that dealers demanded larger amounts (see Appendix Figure A3). Further, since dealers are supposed to bid competitively, and are given a price range when bidding, increasing the total demand decreases the slope in the dealer's bidding function and willingness to pay during the exemption period (relative to the case in which we normalize demand by the supply). The model rationalizes smaller slopes by smaller risk-aversion and shadow cost parameters.

Third, we verify robustness with respect to our measure of volatility. In our benchmark specification, we construct volatility using trades where we observe dealers selling the to-be-auctioned security in a five day window prior to auction. This is natural given that most trading prior to an auction occurs in the one week between the tender open call and the auction close. The more days we include, the larger the volatility. This effect is stronger during the exemption period than during regular times, so that the slope in the dealer's willingness to pay is steeper when assuming zero shadow costs. To rationalize that the observed slope is lower, the shadow costs are higher than in the benchmark specification. This effect goes in the opposite direction when including fewer days to construct the volatility index. Moreover, the fewer days we include, the more likely it becomes that a security is not traded, so that the volatility index is missing for the auction of that security. To avoid dropping these auction entirely, we use the average volatility of same maturity-type auctions within the quarter—in our benchmark specification there is no need to do this.

In addition, we could estimate our model using different volatility indices. One alternative is to use the Implied Volatility Index for Canadian Treasuries, which measures the expected volatility in the Treasury market over the next 30 days (Chang and Feunou (2014)). Given that this volatility drops more strongly during the exemption period than our volatility index, shadow cost estimates are higher when relying on the implied volatility.

Another alternative is to construct return volatility using post-auction trades. We refrain from doing so, because dealers do not know what happens after the auction at the time they bid, so that it seems unnatural to assume that they know the volatility in prices obtained after the auction. Further, post-auction prices likely depend on the realization of the dealer's private information, and with that their willingness to pay, in the auction. This implies that the post-auction volatility—an independent variable in equation (14)—is a function of the dependent variable, and would lead to a simultaneous equation bias.

E Proofs

We first present the proofs of all propositions, and then of all corollaries.

Proof of Proposition 1. We consider the case in which supply has bounded support, $[0, \overline{A}]$, but the proof generalizes to the case of unbounded support. For ease of notation, we omit the type θ_i in this proof, and instead include an $i-$subscript, e.g., $p_i(a) = p(a, \theta_i)$.

Consider dealer i, and fix all other demand schedules at the equilibrium. To determine her best-response, dealer i solves maximization problem (5). To simplify this problem, let $v_i(a) = \frac{\partial V_i(a)}{\partial a}$, denote $p_i'(a) = \frac{\partial p_i(a)}{\partial a}$, and abbreviate all functions, for instance, $p_i(\cdot)$ by p_i. Further, let $\overline{a}_i^c$ be the largest amount that bidder i can win when submitting any demand function given others play an equilibrium function, and $\overline{a}_i^*$ be the largest amount the bidder wins when playing the equilibrium strategy. With this, and auxiliary distribution $G_i(a, p)$ (which is defined in Proposition 1), the dealer's maximization problem becomes:

$$\max_{p_i \in \mathcal{B}} I(p_i) \text{ subject to } L(p_i) \geq 0, \text{ with} \tag{20}$$

$$I_i(p_i) = \int_0^{\overline{A}} F_i(p_i(a), p_i'(a), a) da \text{ with } F_i(p_i(a), p_i'(a), q) = [v_i(a) - p_i(a) - ap_i'(a)][1 - G_i(a, p_i(a))],$$

$$L_i(p_i) = E - \int_0^{\overline{A}} H_i(p_i(a), p_i'(a), a) da \text{ with } H_i(p_i(a), p_i'(a), a) = \kappa[p_i(a) + p_i'(a)a][1 - G_i(a, p_i(a))].$$

Here we have integrated by parts to obtain $I(p_i)$ and $L(p_i)$. A function p_i^* is optimal if the following conditions are satisfied:

$$\frac{\partial(F_i+\lambda_i H_i)}{\partial p_i}(p_i^*(a),p_i^{*'}(a),a)-\frac{d}{da}\left(\frac{\partial(F_i+\lambda_i H_i)}{\partial p_i'}(p_i^*(a),p_i^{*'}(a),a)\right)=0 \text{ for all } a\in[0,\overline{a}_i^*], \quad (21)$$

$$L_i(p_i^*)\geq 0 \text{ and } \lambda_i\geq 0, \quad (22)$$

$$\frac{\partial(F_i+\lambda_i H_i)}{\partial p_i'}(p_i^*(0),p_i^{*'}(0),0)=\frac{\partial(F_i+\lambda_i H_i)}{\partial p_i'}(p_i^*(\overline{a}_i^*),p_i^{*'}(\overline{a}_i^*),\overline{a}_i^*)=0. \quad (23)$$

The last two conditions are the natural boundary conditions. They hold automatically given that $\frac{\partial(F_i+\lambda_i H_i)}{\partial p_i'}(p_i^*(a),p_i^{*'}(a),a)=-(1+\lambda_i\kappa)a[1-G_i(a,p_i^*(a)]$, and $G_i(0,p_i^*(0))=0$, and $G_i(\overline{a}_i^*,p_i^*(\overline{a}_i^*))=1$ by definition of G_i.

Simplifying (21) gives: $-(1+\lambda_i\kappa)[1-G_i(a,p_i^*(a)]-[v_i(a)-(1+\lambda_i\kappa)(p_i^{*'}(a)a+p_i^*(a))]\frac{\partial G_i(a,p_i^*(a)}{\partial p_i}-\frac{d}{da}\big(-(1+\lambda_i\kappa)a[1-G_i(a,p_i^*(a)]\big)=0$, where $\frac{d}{da}\big(a[1-G_i(a,p_i^*(a)]\big)=[1-G(a,p_i^*(a))]-a\big[\frac{\partial G_i(a,p_i^*(a)}{\partial a}+\frac{\partial G_i(a,p_i(a)}{\partial p_i}p_i^{*'}(a)\big]$. This rearranges to condition (6). □

Proof of Proposition 3. The proof of statement (i) is analogous to the proof of Proposition 1. There is only one difference, which comes from the fact that bidders pay the prices they bid for all units that they win instead of the market clearing price. This implies that $I_i(p_i)$ in maximization problem (20) is

$$I_i(p_i)=\int_0^{\overline{A}}F_i(p_i(a),a)da \text{ with } F_i(p_i(a),q)=[v_i(a)-p_i(a)][1-G_i(a,p_i(a))].$$

With slight abuse of notation, we are using the same labels as for the uniform price auction.

The proof of statement (ii) follows from Kastl (2012)'s original proof. The only difference is that the objective function is the Lagrangian, which is analogous to (20). □

Proof of Proposition 4. When supply follows a Generalized Pareto distribution, we can solve for a function that fulfills condition (15) of Proposition 3. For this, we combine the insight that a dealer bids as if her true willingness to pay was $\frac{v(a)}{1+\lambda\kappa}$ for any given $\lambda\geq 0$, with a known result form the literature on equilibrium existence (e.g., Proposition 7 of Ausubel et al. (2014), Theorem 2 of Wittwer (2018))). In equilibrium, $\lambda > 0$ is pinned down by the capital constraint if the constraint binds, and is zero otherwise. □

Proof of Proposition 5. To derive the equilibrium of the proposition, we guess and verify. We guess that there is a symmetric equilibrium in which all contracts are the same, $\phi_i = \phi$, and all trading desks i choose the same demand, $a(p) = \left(\frac{N-2}{N-1}\right)\frac{1}{\rho_I \sigma \phi}(\mu - p)$, for each p, and level of effort, $s_i = 0$. To verify that this equilibrium exists and derive the functional form for ϕ, we begin in the auction stage. We let all trading desks other than i play the symmetric equilibrium and determine trading desk i's best-response in the auction. Then, we find contract ϕ_i that trading desk i's intermediary and households choose assuming that $\phi_j = \phi$ for all $j \neq i$. The proof is complete when we have shown that the best-responses equal the guessed equilibrium.

A trading desk with contract ϕ_i chooses her demand function $a_i(\cdot)$ and whether to exert effort or not, $s_i \in \{0, 1\}$, to maximize the expected utility she obtains from wealth

$$\omega_i(a_i(p), s_i) = \phi_i\{a_i(p)(R - p) - s_i\Delta\} + s_i b \tag{24}$$

point-wise for each p and subject to market clearing, i.e., $\sum_i a_i(p) = A$. If the trading desk exerts effort ($s_i = 0$), her wealth in period 2 is ϕ_i of the return that the asset will generate, which is $a_i(p)(R - p)$. If she shirks ($s_i = 1$), she obtains benefit b but suffers a loss which comes from the fact that the total generated wealth reduces by Δ. Given her contract, the trading desk's loss is ϕ_i of that. Thus, the trading desk exerts effort if the benefit of doing so is larger than the cost, which is the case when

$$\phi_i\Delta \geq b \Leftrightarrow \phi_i(1 + m) \geq 1 \text{ where } m = \frac{\Delta}{b} - 1. \tag{25}$$

Maximizing the objective function point-wise and imposing market clearance, we find that desk i chooses

$$a_i(p) = \left(\frac{\phi \rho_I \sigma}{N-2} + \phi_i \rho_I \sigma \right)^{-1} (\mu - p) \tag{26}$$

in response to all other trading desks choosing the equilibrium guess, and that the auction clears at

$$P^c = \mu - \frac{(N-1)\phi\phi_i \rho_I \sigma A}{(N-2)(\phi + (N-1)\phi_i)}. \tag{27}$$

Anticipating how trading desks behave in the auction, households and the bank choose ϕ_i for trading desk i to maximize their joint expected utility from wealth. Given CARA utility, this is equivalent to

$$\max_{\phi_i} \text{Welfare}(\phi_i) = \sum_{j \in \{H,I\}} a_j(P^c)(\mu - P^c) - \frac{1}{2} a_j^2(P^c) \rho_j \sigma \;\text{ subject to } \phi_i \Delta \geq b, \tag{28}$$

where $a_H(p) = (1 - \phi_i)a_i(p), a_I(p) = \phi_i a_i(p)$, and $a_i(p)$, and P^c are given by (26) and (27), respectively. The solution to this problem pins down a mapping between ϕ_i and ϕ. In the symmetric equilibrium, ϕ_i must equal ϕ. Depending on the size of Δ and b, or equivalently m, there are two solutions to this. For $m \geq 0$, the solution is $\phi_i = \phi = \frac{\Delta}{b}$, or equivalently, $\phi_i = \phi = \frac{1}{1+m}$. For $m < 0$, which is not the case we focus on, the solution is $\phi = 1 + \frac{\rho_H}{(N-2)(\rho_H + \rho_I)}$. Inserting this $\phi_i = \frac{1}{1+m}$ into the market clearing price completes the proof. □

Proof of Proposition 6. Since the moral hazard incentive constraint binds, so that $\phi_i = \frac{1}{1+m}$, whenever the capital constraint binds, the proof is analogous to the proof of Proposition 5. □

Proof of Corollary 1. (*i*) To show that there is no linear equilibrium, we take the perspective of dealer i and fix all other dealers' demand functions. Dealer i chooses an optimal quantity point a for each price p at which the market might clear. The point-wise first-order condition, which is the analogue to conditions (10) an d (12), is

$$\frac{\mu - \rho\sigma[a + z_i]}{1 + \lambda_i \kappa} = p + \mathbb{E}[\boldsymbol{\Lambda_i} | p, \lambda_i][a + z_i], \tag{29}$$

where $\Lambda_i = \frac{\partial p}{\partial a}$ is dealer i's price impact, and λ_i denotes the Lagrange multiplier of her capital constraint. In equilibrium, dealer i's price impact is equal to the slope of his inverse residual supply curves, i.e., for each i, $\Lambda_i = -\left(\sum_{j \neq i} \frac{\partial a_j(\cdot)}{\partial p}\right)^{-1}$. From the first-order condition, we know that dealer i's best-response is linear if and only if $\mathbb{E}[\boldsymbol{\Lambda_i}|p, \lambda_i] = \mathbb{E}[\boldsymbol{\Lambda_i}|\lambda_i]$, or equivalently, observing price realization p does not update the dealer's belief about other dealer's' constraints. However, even if we assume that this holds for all dealers i, the equilibrium price is a function of λ_i of all i. Therefore, there cannot be a linear equilibrium.

(*ii*) Now let $N \to \infty$, so that each dealer's price impact converges to 0, and the market becomes perfectly competitive. Then, the following condition

$$\frac{\mu - \rho\sigma[a + z_i]}{1 + \lambda_i \kappa} = p \Leftrightarrow \tilde{v}_i(a) = (1 + \lambda_i \kappa)^{-1}(\mu - \rho\sigma(a + z_i)) \tag{30}$$

characterizes the equilibrium demand of dealer i. This equilibrium exists if there are $\lambda_i \geq 0$ for all i such that the capital constraints are satisfied. The market clears at

$$P^{\infty} = \lim_{N \to \infty} \frac{\mu - \sigma\rho(\frac{A}{N} + \frac{1}{N}\sum_i z_i)}{1 + \frac{1}{N}\sum_i \lambda_i \kappa} = \frac{\mu - \sigma\rho\mathbb{E}[\boldsymbol{z_i}]}{1 + \mathbb{E}[\boldsymbol{\lambda_i}]\kappa} \tag{31}$$

as $\lim_{N\to\infty} \frac{1}{N}\sum_i z_i = \mathbb{E}[\boldsymbol{z_i}]$ and $\lim_{N\to\infty} \frac{1}{N}\sum_i \lambda_i = \mathbb{E}[\boldsymbol{\lambda_i}]$ by the law or large numbers. □

Proof of Corollaries 2 and 3. (*i*) When dealers only face aggregate uncertainty, equilibrium demand is given by Proposition 2. Further, we can infer the price impact $\Lambda = \frac{1}{N-1}\frac{\sigma\rho}{1+\lambda\kappa}$ and the market clearing price: $P^*(\Lambda) = \frac{1}{1+\lambda\kappa}\left(\frac{1}{N}\sum_i \mu_i - \frac{N-1}{N-2}\frac{A}{N}\sigma\rho\right)$, with $\mu_i = \mu - \sigma\rho z_i$. In contrast, when bidders are price-takers and submit their true willingness to pay the market clears at: $P^*(0) = \frac{1}{1+\lambda\kappa}\left(\frac{1}{N}\sum_i \mu_i - \frac{A}{N}\sigma\rho\right)$. Thus, $markup = P^*(0) - P^*(\Lambda) = \frac{1}{1+\lambda\kappa}\frac{A}{N}\frac{1}{N-2}\sigma\rho$.

From here it is easy to see that the slope of equilibrium demand (9), the market price $P^*(\Lambda)$, price impact and the markup increase when $\lambda\kappa$ decreases. Further, we can compute the following elasticity

$$\eta = \frac{\partial markup}{\partial \lambda\kappa}\frac{\lambda\kappa}{markup} = \frac{P^*(\Lambda)}{\partial \lambda\kappa}\frac{\lambda\kappa}{P^*(\Lambda)} = \frac{\partial \Lambda}{\partial \lambda\kappa}\frac{\lambda\kappa}{\Lambda} = \frac{1}{1+\lambda\kappa} - 1. \tag{32}$$

(ii) Now consider the case in which dealers have private information, so that Proposition 1 applies. When the market is perfectly competitive and all dealers are price-takers, dealer demand is given by $\tilde{v}_i(a) = (1+\lambda_i\kappa)^{-1}(\mu - \rho\sigma(a+z_i))$ according to Corollary 1. The market clears at the price given in (31). When shadow costs decrease for all i, the demand curve becomes steeper and the market price increases. Further, we can compute elasticity η^∞ analogously to before. When at least some dealers share their bids, the effects depend on the way the shading factor changes relative to the true willingness to pay. This, in turn, depends on the distribution of signals, and supply, and the number of competing dealers. □

Proof of Corollary 4. To show that there is no $\phi_i = \phi$ for all i that implements the price and allocation of a frictionless market, we compute the price and allocation of such a frictionless market and compare both to the analogue in our market setting.

In the frictionless market, agent of type $j \in \{H, I\}$ submits the following demand $\bar{a}_j(p) = \frac{1}{\rho_j\sigma}(\mu - p)$. Intuitively, each agent submits the marginal utility she achieves from winning amount $a_j(p)$, conditional on the auction clearing at p. The market would clear at price $\bar{P}^c = \mu - (\rho_H + \rho_I)\sigma\frac{A}{N}$ at which $N\bar{a}_H(\bar{P}^c) + N\bar{a}_I(\bar{P}^c) = A$. Households obtains $\bar{a}_H(\bar{P}^c) = \frac{A}{N}\frac{\rho_H+\rho_I}{\rho_H}$; the bank obtains $\bar{a}_I(\bar{P}^c) = \frac{A}{N}\frac{\rho_H+\rho_I}{\rho_I}$.

Comparing this price and the allocation to the one presented in Proposition 5, we see that it is not possible to obtain the frictionless price and frictionless allocation with the same ϕ. We can only obtain one of the two. □

Proof of Corollary 5. Statement (i) follows from Proposition 5 and Corollary 4. To show statement (ii) we only need to determine how $\lambda > 0$ changes in N, since we already know that the market price increases when λ decreases.

$$\frac{\partial \lambda}{\partial N} = \frac{A(-\mu(N-2)^2 N + (4 + N(2N-5))\phi\rho_I\sigma A)}{E(N-2)^2 N^3}$$

$$\frac{\partial \lambda}{\partial N} \begin{cases} < 0 & \text{if } \mu > c(N) = \frac{(4+N(2N-5))\phi\rho_I\sigma A}{(N-2)^2 N} \\ > 0 & \text{otherwise} \end{cases}$$

Note that cutoff $c(N)$ strictly decreases in N, and converges to 0 as $N \to \infty$. Therefore, given $\mu > 0$, there is some $\bar{N}$ at which $\mu = c(\bar{N})$ so that for $N > \bar{N}$, $\frac{\partial \lambda}{\partial N} < 0$ and for $N < \bar{N}$, $\frac{\partial \lambda}{\partial N} > 0$. If $\mu > \frac{7}{3}\rho_I\sigma\phi A$, $\frac{\partial \lambda}{\partial N} < 0$ for any N. □

Appendix Table A1: Bid functions are approximately linear

	mean	median	sd
β_t	0.21	0.16	0.13
R_t^2	0.72	0.74	0.11
Adj. R_t^2	0.64	0.67	0.15
Within R_t^2	0.54	0.56	0.15

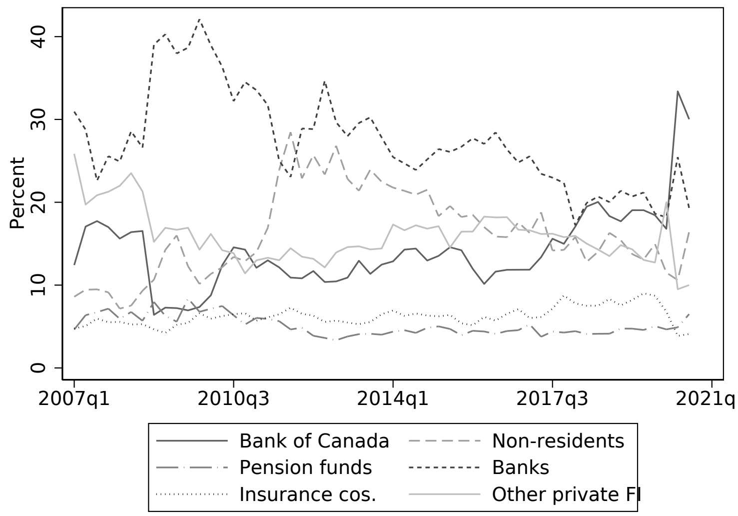
Appendix Table A1 shows the point estimate and R^2 from regressing bids on quantities in each auction: $b_{tik} = \zeta_{ti} + \beta_t a_{tik} + \epsilon_{tik}$. The subsample are bidding-functions with at least 2 steps. Bids are in yields (bps) and quantities in percentage of supply.

Appendix Table A2: Robustness w.r.t. risk aversion

	(All)		(2020)		(2022)	
Main specification	0.0065	(0.0051)	0.0065	(0.0038)	0.0074	(0.0070)
Number of steps ≥ 3	0.0067	(0.0051)	0.0065	(0.0039)	0.0076	(0.0073)
Number of steps ≥ 4	0.0066	(0.0051)	0.0063	(0.0037)	0.0077	(0.0073)
Number of steps ≥ 5	0.0066	(0.0052)	0.0058	(0.0034)	0.0076	(0.0074)
Number of steps ≥ 6	0.0061	(0.0056)	0.0046	(0.0030)	0.0086	(0.0081)
Quantities in mil C$	0.0002	(0.0001)	0.0001	(0.0001)	0.0002	(0.0002)
Volatility using 1 day	0.0063	(0.0058)	0.0059	(0.0044)	0.0067	(0.0070)
Volatility using 2 days	0.0056	(0.0054)	0.0046	(0.0037)	0.0067	(0.0069)
Volatility using 3 days	0.0061	(0.0057)	0.0056	(0.0042)	0.0066	(0.0069)
Volatility using 4 days	0.0066	(0.0053)	0.0064	(0.0041)	0.0073	(0.0071)
Volatility suing 6 days	0.0066	(0.0052)	0.0066	(0.0041)	0.0074	(0.0065)
Volatility using 7 days	0.0067	(0.0052)	0.0067	(0.0047)	0.0071	(0.0060)
Volatility using 8 days	0.0069	(0.0049)	0.0067	(0.0046)	0.0069	(0.0059)
Volatility using 9 days	0.0068	(0.0048)	0.0067	(0.0046)	0.0068	(0.0058)
Volatility using 10 days	0.0071	(0.0049)	0.0071	(0.0049)	0.0067	(0.0055)

Appendix Table A2 presents the median of all risk-aversion estimates for all specifications in column (All), the median of estimates around the exemption in column (2020) and when the exemption ended in column (2022). The main specification uses functions with at least 2 steps, expresses quantities in percentage of supply, and relies on the volatility index that uses trades during 5 business days before the day of the auction. The second to fifth row show robustness with respect to the number of steps of the bidding/willingness to pay functions. The sixth row presents results when using quantities in million C$. The remaining rows display the results for different volatility indices, constructed using N trading days prior to the auction, for N=1, ..., 10. Standard errors are presented in parentheses. They are stated in multiples of 100 to reduce the number of zeros.

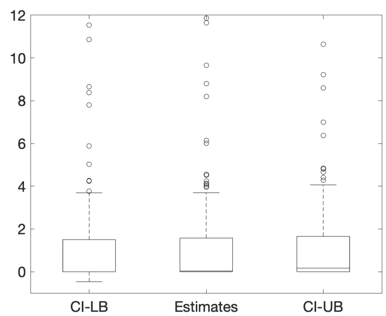
Appendix Figure A1: Holders of Canadian government bonds



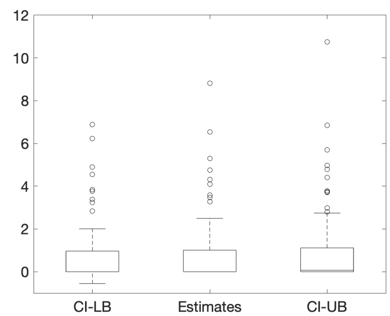
Appendix Figure A1 shows who holds Canadian government bonds and bills from 2007 until 2021 in percentage of par value outstanding: Bank of Canada, Non-residents, Canadian pension funds, Canadian banks, Canadian insurance companies, and other private firms. The bank category holdings are mostly driven by the eight banks we focus on. They hold over 80% of the assets.

Appendix Figure A2: Robustness w.r.t. shadow costs

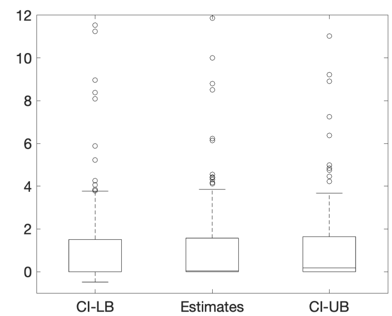
(a) Main specification



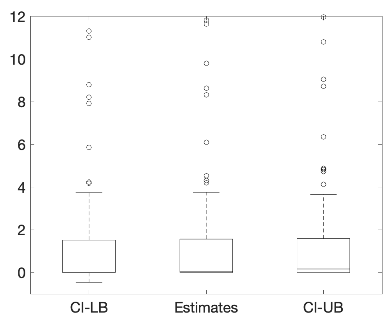
(b) With quantities



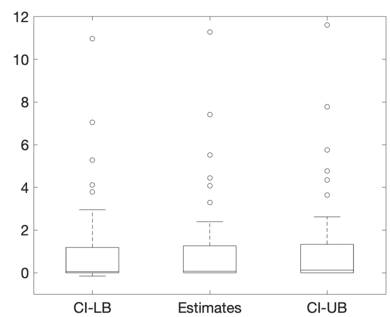
(c) Number of steps ≥ 3



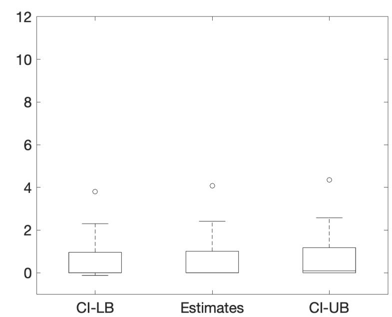
(d) Number of steps ≥ 4



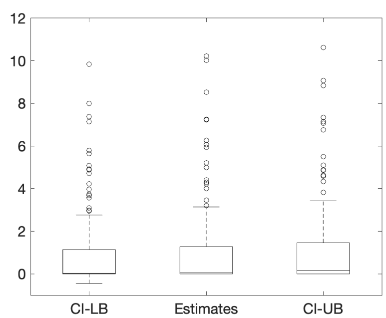
(e) Number of steps ≥ 5



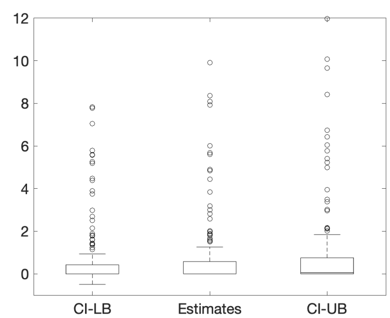
(f) Number of steps ≥ 6



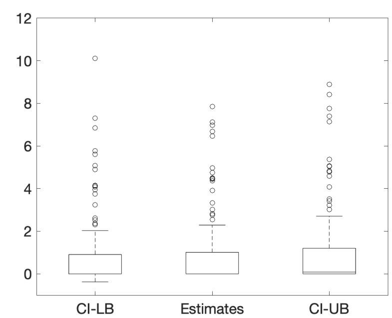
(g) Volatility (1 days)



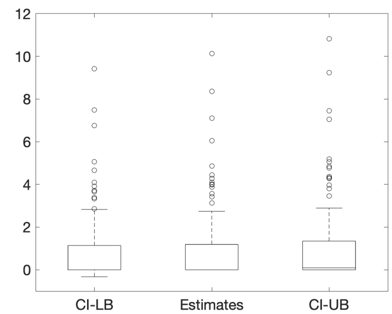
(h) Volatility (2 days)



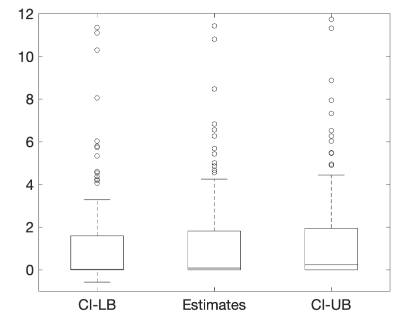
(i) Volatility (3 days)



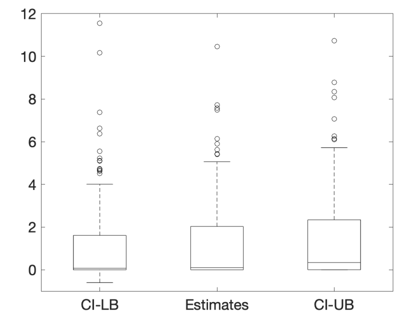
(j) Volatility (4 days)



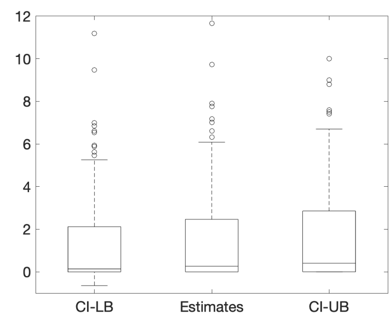
(k) Volatility (6 days)



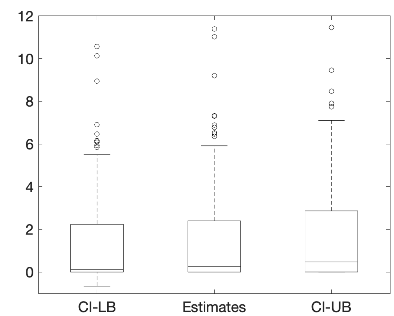
(l) Volatility (7 days)



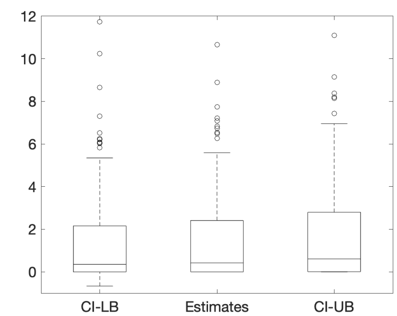
(m) Volatility (8 days)



(n) Volatility (9 days)



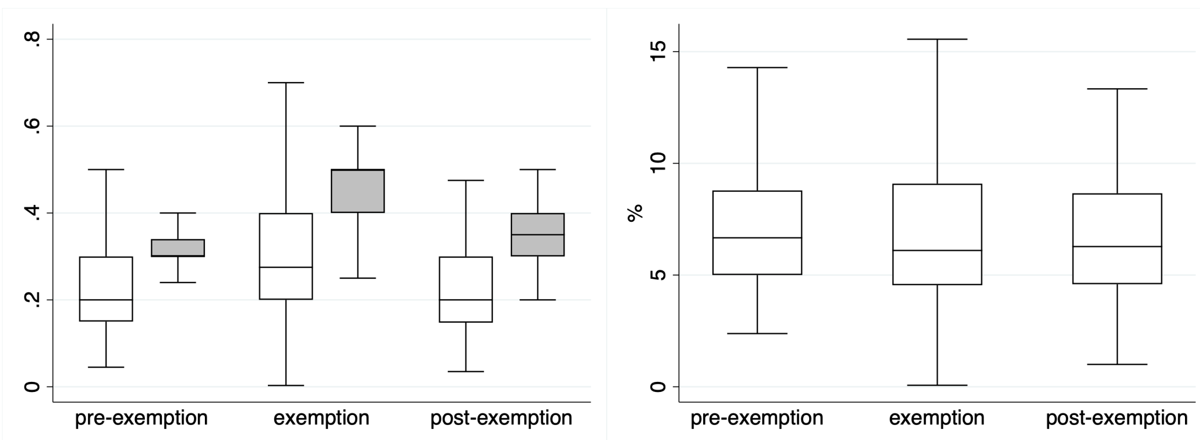
(o) Volatility (10 days)



Appendix Figure A2 shows the distribution of the lower and upper bounds of the 95% confidence intervals for each point-estimate (CI-LB and CI-UB, respectively), in addition to the distribution of the point estimates (Estimates) pooled across all auctions, and excluding outliers for all model specifications. Panel (a) is identical to the RHS of Figure 7. In (b) we use quantities in million C$, in (c)-(f) we change the number of steps of value functions that are included in the estimation, in (g)-(o) we use different volatility indices, that are constructed pooling trades during 1-10 business before the auction takes place.

Appendix Figure A3: Variation in quantities

(a) Supply and total demand (b) Total demand in percentage of supply



Appendix Figure A3a shows the distribution of the total amount a dealer demands in an auction before, during and after the exemption period (in white), and the distribution of the supply (in gray). Demand is expressed in million C$, and supply is in 10 million C$ to make the two comparable. Appendix Figure A3b shows the distribution of the total amount demanded as percentage of supply across periods.