NORTHWESTERN UNIVERSITY

Essays in Financial Economics

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Financial Economics

By

Jingxiong Hu

EVANSTON, ILLINOIS

June 2023

ABSTRACT

Essays in Financial Economics

Jingxiong Hu

In Chapter 1, I analyze optimal capital structure using a model in which firms issue securities in order to (1) finance investments in operations and (2) recapitalize the firm. In this trade-off model, firms balance the tax benefits of debt against the costs of financial distress. Key to the analysis, the marginal tax benefit of debt depends on whether the debt is used for financing investments or financial restructuring. The theory explores leverage dynamics with personal and corporate taxes in a trade-off model with continuous leverage adjustments and no security issuance costs. Depending on firms' external financing needs, the marginal source of financing can be debt or equity. There are two local leverage targets for firms with leverage above or below a threshold. The model generates a leverage distribution that closely matches the data, including many zero-leverage firms. Policymakers can reduce the expected bankruptcy loss without losing tax revenue by taxing shareholders more at the personal level and less at the corporate level.

In Chapter 2, I estimate a new measure of marginal tax "benefits" of debt issuance that accounts for personal taxes from a dynamic perspective. Previous literature overestimates

the marginal tax benefits of debt by definition since they implicitly assume that debt issuance always leads to a reduction in equity value. However, debt issuance only reduces equity value when replacing equity issuance. If the proceeds from debt issuance are expected to generate additional payouts, net personal tax costs can exceed the corporate tax savings of debt issuance, leading to a negative tax benefit even if the firm has positive taxable earnings. I find that there is no marginal benefit to borrowing for seemingly underleveraged U.S. firms. The new measure explains the observed leverage changes better than the traditional measure.

In Chapter 3, we study the microstructure of the financial market for close trading and the effects on price informativeness. Passive investment strategies that trade at market close have incurred high transaction fees charged by the primary exchanges. Investment banks undercut the exchanges by executing client orders at close prices set on the exchanges yet charging lower fees. While providing liquidity, banks trade on the order flow information. Using a quasi-experimental shock – an NYSE close auction fee cut – we find that banks' trading activities improve the informativeness of close prices and reduce the cost of passive investment strategies. To explain this finding, we propose a model where dual trading improves price discovery. A bank contributes to price discovery by trading on the informativeness of the orders it receives relative to the market. The implications of our model apply generally to scenarios with multiple trading venues where venue operators trade on order flow data. As an application of our results, the common practice of trading on retail order flow information could potentially improve price discovery.

Acknowledgements

I am deeply indebted to my committee members, Michael Fishman (chair), Nicolas Crouzet, Zhiguo He, Konstantin Milbradt, and Mitchell Petersen. I would never have accomplished this without their continuous guidance and support.

For Chapters 1 and 2, I also thank Robert McDonald, Matthew Rognlie, Fabrice Tourre, and all the faculties who attended my presentations at Northwestern University for their comments and feedback.

Chapter 3 was coauthored with Jiaheng Yu. This paper supersedes our earlier work, "Dual Trading and Price Discovery at Market Close: Theory and Evidence," joint with Canyao Liu. We thank, in addition, Nicholas Barberis, Alexandru Barbu, Sabrina Buti, Hui Chen, Yixin Chen, James Choi, Piotr Dworczak, Thomas Ernst, Stefano Giglio, Travis Johnson, Robert Korajczyk, Eben Lazarus, Joshua Mollner, Zhongling Danny Qin, J.R. Scott, Kelly Shue, Jian Sun, Yupeng Wang, Liyan Yang and Haoxiang Zhu, as well as seminar participants at Midwest Finance Association 2020, MIT Sloan, Northwestern Kellogg, Trans-Atlantic Doctoral Conference 2021, Yale SOM, Wharton Inter-Finance PhD Seminar, AsianFA 2022, EuroFA 2022, NFA 2022, and FMA 2022 for invaluable comments and suggestions.

Finally, I especially thank my beloved parents and wife for their unwavering support and encouragement.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

A dynamic trade-off theory with corporate and personal taxes

1.1. Introduction

The tax shield of debt is pivotal for firms' capital structure policies. The trade-off theory of capital structure suggests that firms should lever up until the marginal cost of incremental bankruptcy risk offsets the marginal tax benefit. On average, however, firms shield less than 1/3 of their earnings with interest expenses (Berk and DeMarzo, 2011, Chapter 15). Furthermore, about 1/5 of firms have close to zero leverage (Strebulaev and Yang, 2013).[1] These low-leverage firms typically have high profits and good liquidity, which suggests a low risk of financial distress. It is, therefore, puzzling that these firms do not lever up to take advantage of the tax shield.

This paper demonstrates that, when properly considered, corporate and personal taxes themselves—not other frictions—can explain this puzzle. The literature and most textbooks traditionally interpret the tax shield of debt using Miller's (1977) definition, $(1-\tau_b)-(1-\tau_c)(1-\tau_e)$, which evaluates the tax benefits of firms' outstanding debt by comparing the difference between the tax costs of delivering 1 dollar to equity holders and delivering 1 dollar to debtholders. When firms pay debtholders, the debtholders are charged a personal tax on income from bonds at rate τ_b. When firms pay equity holders, there is a corporate income tax at rate τ_c that is charged on firms' profits, as well as a

[1] The fraction of firms with less than 5% book leverage increases to about 1/4 when we extend the sample period to 1962-2021, using the same definition as in Strebulaev and Yang (2013), which documents the fact for 1962-2009.

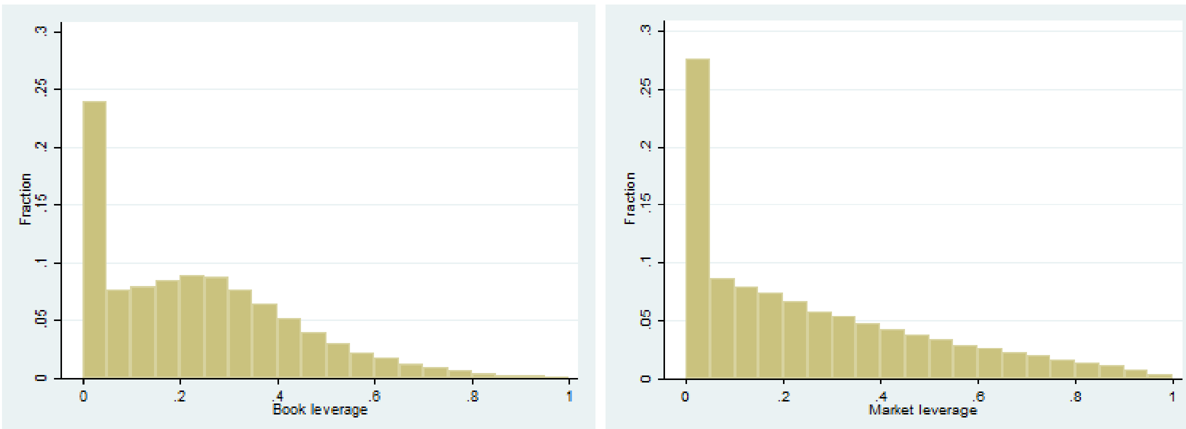
personal tax on income from equity at rate τ_e that is charged on equity holders' capital gains. This definition only counts taxes on cash flows in the future and may not serve as a proper assessment of the tax benefits of issuing new debt because it fails to capture the tax consequence of the proceeds from debt. As the proceeds from debt are available to shareholders and added to the firm's value, they are subject to a personal tax on income from equity τ_e, either in the form of a dividend tax if they are directly distributed as dividends or in the form of a capital gains tax otherwise.

Following this idea, I examine a trade-off theory and its empirical implications. I develop a dynamic capital structure model with continuous leverage adjustments and no fixed costs for issuing securities following DeMarzo and He (2021). The key frictions are just corporate and personal taxes and bankruptcy costs. The model generates leverage dynamics in which capital structure policies depend on firms' financing needs, and the marginal source of financing can be debt or equity. As a result, a firm's leverage slowly adjusts to one target or another when leverage is above and below a threshold and may switch targets if leverage crosses the threshold due to shocks. With reasonable parameter choices, the model's simulated leverage distribution closely matches the leverage distribution of Compustat firms, with zero-leverage firms representing about 1/4 of all firms. The model produces novel implications for tax policies. Policymakers can reduce the ex-ante expected bankruptcy loss due to leverage distortions from taxes without losing tax revenue by charging a lower corporate tax rate and a higher personal tax rate on income from equity.

My theory propose a solution to a long-lasting puzzle in the literature that firms seem to exploit the tax shield of debt inadequately. Graham (2000) notes low-leverage firms' puzzling choices not to lever up in his conclusion: "Paradoxically, large, liquid, profitable firms with low expected distress costs use debt conservatively." And Strebulaev and Yang (2013) demonstrate that trade-off models of capital structure in the previous literature (e.g., Goldstein, Ju, and Leland, 2001; Strebulaev, 2007) hardly account for the cross-sectional distribution of corporate leverage ratios, especially the presence of numerous firms with less than 5% leverage. Figure 1.1 shows the cross-sectional distribution of book and market leverage among Compustat nonfinancial firms in the U.S. from 1962 to 2021. The figure demonstrates that about 1/4 of the firm-year observations have leverage below 5%. Furthermore, the fractions of observations within each 5% bin of market leverage ratios decrease in leverage levels. It is not surprising that standard trade-off models with a single positive leverage target cannot generate this type of distribution. Given the leverage target, we would expect a bell-shaped cross-sectional distribution that may be skewed and tops at the leverage target.

Figure 1.1. Leverage of Compustat nonfinancial firms in the US, 1962-2021

(a) Book leverage (b) Market leverage



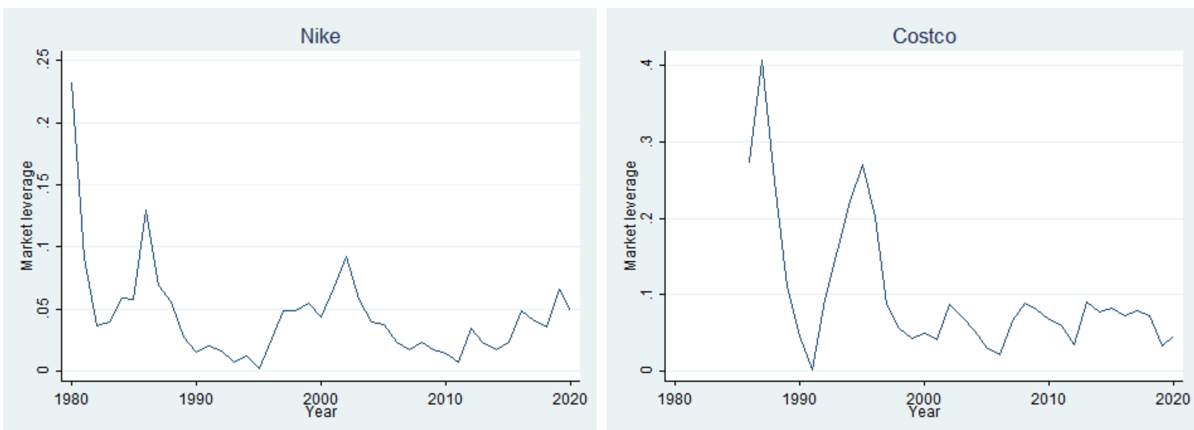
Notes. Panel (a) plots the book leverage, defined by $\frac{\text{Book value of debt } (DLTT+DLC)}{\text{Book value of asset } (AT)}$, of firms headquartered in the U.S. in the Compustat-CRSP merged data set from 1962 to 2021 annually. Firms in the financial industry [Standard Industrial Classification (SIC) codes 6000-6999], utilities (SIC codes 4900-4999), American depository receipts ("ADR") (SIC codes 8800-8999), non-publicly traded firms [stock ownership variable (STKO) 1 or 2], and firm-years with total book value of assets (AT) less than 10 million inflation-adjusted year 2000 dollars are excluded. There are 283,702 firm-year observations. Panel (b) plots the market leverage, defined by $\frac{\text{Book value of debt } (DLTT+DLC)}{\text{Book value of debt } (DLTT+DLC)+\text{Market value of equity } (CSHO\times PRCC\_F)}$ of these firms.

DeMarzo and He (2021) explain the puzzle by a commitment problem: investors expect firms to maintain high leverage and charge a high credit spread so that firms gain no tax benefit from debt. In their model, firms either never issue debt or continuously issue debt. In the data, however, many low-leverage firms had greater leverage in their earlier years. For example, Nikes and Costco's market leverages, as seen in Figure 1.2, were higher than 20% and 40%, respectively, in the 1980s, but they were lower than 5% in recent years.

[2] In addition, firms that slowly adjust to a high leverage target, as in their model, are unlikely to generate the shape of the fractions of observations in Figure 1.1 that remain after excluding the zero-leverage firms. A fully satisfactory explanation has not yet been found.

Figure 1.2. Time series of Market leverage for Nike and Costco

(a) Nike (b) Costco



Notes. Panel (a) and (b) plot the time series of Nike and Costco's market leverage at the annual frequency, defined by book value of debt / (book value of debt + market value of equity), from their first observation in Compustat data to 2020, respectively.

This paper argues that firms may keep their leverage as low as zero—not due to the high costs of debt or precautionary motives, but because they face no tax benefits from issuing debt. Suppose, for example, that a low-leverage firm attempts to shield its earnings with more interest expenses (net of interest income). In this case, the firm recapitalizes by issuing debt and distributing the proceeds from the debt to shareholders as payouts.

[2]Among firms with less than 5% leverage, 2/3 had more than 10% leverage, and half had more than 20% leverage. Such deleveraging is suboptimal for firms in the DeMarzo and He (2021) model, as it leads to a value transfer from shareholders to debtholders.

Shareholders must pay dividend taxes immediately if the proceeds are distributed as dividends. Otherwise, the share repurchase increases the stock value, thereby increasing the capital gains taxes that shareholders pay. Assuming that the firm's payout policy is constant over time, this incremental personal tax for shareholders fully offsets their personal tax savings from interest expenses. Indeed, debt issuance does not directly result in savings on personal tax on income from equity for shareholders; rather, it only transfers it intertemporally. The tax savings (or costs) for such recapitalization, then, depends on the comparison between the firm's corporate tax rate and the bondholders' personal income tax rate. Since the top federal personal tax rates are higher than the top corporate tax rates in most years, it is reasonable that firms have little or no tax incentive for such recapitalization.[3]

I develop a continuous-time trade-off model with corporate and personal taxes in which the firm can continuously adjust its leverage, as in DeMarzo and He (2021). I solve a smooth equilibrium without discrete leverage adjustments. Equity value and debt price functions are determined by a pair of piecewise non-linear non-homogenous differential equations that have no known closed-form solutions. This paper, therefore, develops a novel numerical method based on a fourth-order Runge-Kutta-Nystrm algorithm to solve the differential equations. The results of the model are expressed directly as functions of the equity value and debt price functions. These analytical expressions deliver interesting results even before the model is fully solved.

The tax benefits (or costs) of issuing a marginal dollar of debt can be decomposed into two parts. First, each dollar of the firm's EBIT is either taxed by the corporate rate or the

[3]The empirical literature (e.g., Ang, Bhansali, and Xing, 2010; Longstaff, 2011) finds that the implicit tax rates priced in bonds are close to the top federal tax rates.

personal rate on income from bonds, depending on whether or not it is used for interest expenses. The firm, thus, saves the difference between the two rates with each dollar of interest payment. Second, debt issuance leads to changes in current and future payouts. The firm saves the personal rate on equity income with each dollar reduction in net payouts. This part is negative when the proceeds from debt are distributed as payouts and positive otherwise. With reasonable corporate and personal tax rates values, a marginal debt issuance brings the firm positive or negative tax savings depending on its current and expected future financing margins. When the firm's external financing need is monotone in its leverage, I characterize its optimal financing policies, depending on parameters, by at most four financing rules for different regions of state variables. The marginal source of financing in each region can be equity issuance, debt issuance/repurchase, or dividend distribution.

The model generates leverage dynamics with two local leverage targets for firms with higher and lower leverage than a threshold level. When the corporate tax rate is not higher than the personal tax rate for bond income, high-leverage firms slowly adjust to a high leverage target, and low-leverage firms slowly adjust to zero leverage. Since firms with higher leverage need more external financing due to payments to debtholders, leverage differences between firms persist even if they have identical earnings flow afterward. A simulation of the model generates a stationary cross-sectional leverage distribution that closely matches the leverage distribution in the data, in which about 1/4 of firms have close to zero leverage.

The model has several implications for economic efficiency. First, it allows us to determine the levels at which a policymaker should set the tax rates. This conclusion is possible because the corporate and personal taxes on equity income in the model affect the firm's financing policies differently—in contrast to Miller's formula in which they are always charged together. I decompose the firm's pre-tax value into the values of equity, debt, tax revenue, and expected bankruptcy loss. I find that when a government aims to collect a target level of tax revenue and reduce expected deadweight bankruptcy loss due to leverage distortions, it can improve efficiency by taxing equity holders more at the personal level and less at the corporate level. The firm prefers to issue debt with a longer maturity, but a shorter maturity improves welfare.[4] In an extension with endogenous investment and debt overhang, a corporate tax cut increases investments while a dividend tax cut decreases investments in the short run.[5]

The remainder of the paper is organized as follows. Section 2 reviews the literature. Section 3 discusses potential differences between the tax consequences of debt issuance and changes in the value of tax shield. Section 4 sets up the model and characterizes optimal financing policies. Section 5 demonstrates the leverage dynamics and leverage distribution generated by the model. Section 6 discusses the model's welfare implications. Section 7 concludes.

[4]Long maturity reduces rollover risk as in He and Xiong (2012) and Diamond and He (2014). As a result, the firm takes higher leverage and assumes more bankruptcy risk. In DeMarzo and He (2021), firms are indifferent to longer and shorter maturity since they gain no benefit from debt. A force that favors shorter maturity, for example, could be investors' liquidity preference (He and Milbradt, 2014).

[5]The short-run effects are characterized by investment changes without change in leverage since leverage adjustments are slow. In the long run, however, policies also affect investments through firms' leverage changes. Decreasing the tax benefits of debt reduces debt usage, alleviates debt overhang, and improves investment. DeMarzo and He (2021) and Crouzet and Tourre (2021) show that policies cutting the cost of debt may reduce investment in the long run.

1.2. Literature Review

This paper contributes primarily to two strands of literature. First, it adds to the extensive literature on trade-off models of capital structure. Ai, Frank, and Sanati (2020) provide a thorough review of this literature. These models typically assume a constant rate of tax benefits on interest payments, which can be interpreted as the corporate tax rate or the rate of tax benefits defined by Miller (1977).[6] Exceptions include models with real investment opportunities that incorporate a wide range of frictions (e.g., Hennessy and Whited, 2005, 2007; Gamba and Triantis, 2008) and models that capture the deferral of capital gains tax (Lewellen and Lewellen, 2006; Bolton, Chen, and Wang, 2014).

My consideration of different tax benefits from debt issuance for financing investments and leveraged restructuring is closely related to Hennessy and Whited (2005), which points out the importance of analyzing the trade-off theory dynamically. In their discrete-time model with one-period debt, Hennessy and Whited (2005) show that marginal debt issuance generates more tax savings when replacing equity issuance than financing distributions. Compared to their quantitative model, my model highlights the effect of taxes on leverage dynamics in a framework without any frictions other than taxes and bankruptcy costs. Taxes can generate the observed leverage distribution without fixed costs of security issuance.[7] Such a neat setting is more suitable for studying the implications for tax policies. The continuous-time model with long-term debt allows me to analyze

[6]For example, Leland (1994a, 1994b, 1998), Goldstein, Ju, and Leland (2001), Titman and Tsyplakov (2007), and DeMarzo and He (2021).

[7]Hennessy and Whited (2005) assume the top corporate tax rate to be higher than the personal tax rate on bond income. In their model, firms have low leverage mainly because of precautionary incentives. As a result, their model is unlikely to match the fraction of zero-leverage firms with a reasonable equity flotation cost.

more characteristics of capital structure policies, such as leverage adjustments, leverage targets, and optimal maturity.

In contrast to the standard trade-off theory, since Modigliani and Miller (1963), in which taxes always incentivize firms to take more debt, I argue that the net tax consequence of debt issuance can be a cost instead of a benefit for firms. That is because, in a leveraged recapitalization, tax costs on distributing the proceeds from debt as payouts can exceed the tax savings from debt on firms' future cash flows. Ivanov, Pettit, and Whited (2020) also model a nonstandard relationship between tax rates and leverage usage from a different channel to rationalize their empirical findings that small private firms' leverage rises after tax cuts. In their model, higher tax rates raise the default threshold since defaults are triggered when firms cannot repay their debt by after-tax earnings, thus increasing the cost of debt.

The setting of my model follows a large group of continuous-time dynamic models pioneered by, for example, Fischer et al. (1989), Leland (1994), and Leland and Toft (1996). My model is closest to DeMarzo and He (2021), in which firms can adjust capital structure dynamically at no cost.[8] DeMarzo and He (2021) show that the increase in credit spread caused by the leverage ratchet effect can fully offset the tax benefits of debt when firms cannot commit to a leverage policy.[9] My model introduces personal taxes—and, most importantly, the personal taxation difference between positive and negative payouts—to the DeMarzo and He (2021) framework. Leverage dynamics in my model

[8]Earlier models assume firms never restructure (e.g., Leland, 1994a, 1994b) or retire all debt when restructuring (e.g., Fischer, Heinkel, and Zechner, 1989; Goldstein, Ju, and Leland, 2001; Strebulaev, 2007; Dangl and Zechner, 2021).

[9]See, for instance, He and Milbradt (2016), Admati et al. (2018), and Demarzo (2019) for discussions of the commitment problem.

differ from those in the previous literature, such as DeMarzo and He (2021), in several ways. First, in my model, there are two local leverage targets for firms with higher or lower leverage compared to a threshold, while a standard trade-off theory has a single leverage target. Second, the marginal source of financing can be debt or equity. And third, firms may repurchase debt at the cost of the leverage ratchet effect and gain from the tax shield even without a commitment device.

The simulated leverage distribution of the model can closely match the cross-sectional leverage distribution in the data, in which about 1/5 of firms have close to zero leverage. Explaining the cross-sectional distribution of firms' leverage—and especially the levels of low- and zero-leverage firms—has historically posed a critical challenge to the trade-off theory. The benchmark Leland (1994) model, in which firms cannot adjust their debt levels as earnings grow, predicts a leverage ratio of over 70% with reasonable values of parameters, which is way too high compared to the average market leverage of 26% for Compustat firms in the period between 1987 and 2003 (Strebulaev and Yang, 2013). Models with infrequent leverage adjustments, such as Goldstein et al. (2001), Ju et al. (2005), and Strebulaev (2007), generate more reasonable average leverage ratios, but they generate very few or no firms with leverage below 5%. Finally, models with endogenous investment and fixed costs (e.g., Hennessy and Whited, 2005; Hackbarth and Mauer, 2012; Kurshev and Strebulaev, 2015) can generate zero-leverage firms, but they are unlikely to generate a large proportion of zero-leverage firms with a low cost of debt, as documented in Strebulaev and Yang (2013).

Fixed equity issuance costs can also generate asymmetric benefits of debt with positive and negative payouts (e.g., Cooley and Quadrini, 2001; Hennessy and Whited, 2005;

Kurshev and Strebulaev, 2015; Bolton, Wang, and Yang, 2021). I show that because the personal tax on equity income can only be saved by reducing equity issuance, it effectively impacts firms' capital structure policies through the same mechanism as a cost on equity issuance. The primary difference is in their effects on the size of tax benefits. In addition to the benchmark tax benefits formula in Miller (1977), equity flotation costs increase the benefits of debt when it replaces equity issuance, while a personal tax on payouts from restructuring decreases the benefits of debt when it finances payouts. With reasonable tax rates, then, costs on payouts can better rationalize zero-leverage firms decisions not to recapitalize than costs on equity issuance.

Second, this paper relates to the public economic literature on payout taxation. Early works in the "trapped equity" view (also often referred to as the "new" view) of dividend taxation (e.g., King 1974, 1977; Auerbach, 1979, 1981) point out that dividend taxation does not affect firms' decisions when the firm issues no equity.[10] Internal equity is "trapped" in firms and has to be taxed when distributed to shareholders. Auerbach (2002) reviews literature on this topic. This paper applies a similar argument regarding the tax benefits of debt: debt issuance cannot save tax on shareholders' equity income unless it reduces equity issuance. Relatedly, the paper also contributes to the literature on optimal taxation design. For example, Dávila and Hébert (2019) show that governments should tax financially unconstrained firms to maximize the efficiency of investments and production while collecting a given amount of tax revenue. The optimal policy is to tax firms on their payouts to shareholders instead of their income. This paper achieves the same result in a different setting. Besides the investment channel as in Dávila and Hébert

[10]See, e.g., Poterba and Summers (1984) for discussions of the "traditional" and "new" views of dividend taxation.

(2019), I reveal a new channel working through distortions on firms' leverage that makes payout taxation better than corporate income taxation. While raising the same amount of tax revenue, taxes on corporate income lead to more expected deadweight bankruptcy loss than taxes on payouts to shareholders.

1.3. Tax benefits from debt issuance

Miller (1977) defines the value of tax shield from each dollar (market value) of outstanding perpetual debt by $\left[1 - \frac{(1-\tau_c)(1-\tau_e)}{1-\tau_b}\right]$, which is widely used in the literature and textbooks for evaluating firms' existing debts and the benefits from new debt issuances.[11][12] Changes in the value of tax shields on future earnings, however, may not be the correct measure of tax benefits generated by issuing new debt since they do not capture the potential tax consequences at the debt issuance time. For example, if proceeds from debt issuance are distributed as payouts, taxes on these payouts are not captured by the definition above. Below, I discuss tax incentives for real-world scenarios in which firms may issue new debt, and I consider whether Miller's formula applies in these scenarios.

When considering new debt issuance, a firm may face two scenarios depending on whether an investment opportunity needs to be financed externally and has a positive net present value (NPV) if financed most cheaply. I define the tax benefit as the tax savings by issuing a marginal dollar of debt relative to the best alternative without such debt issuance. In the first case, when there is a positive NPV project to invest in, additional

[11] A scaled definition $(1-\tau_b) - (1-\tau_c)(1-\tau_e)$ describes the difference between after-tax income earned by debtholders and shareholders from each dollar of the firm's earnings. It can be viewed as the tax shield from each dollar of interest expense.

[12] Miller (1977) describes a market equilibrium in which $(1-\tau_b) = (1-\tau_c)(1-\tau_e)$ and there is no optimal capital structure for individual firms. In appendix D, I discuss how the different tax consequences of financing investments by debt and leveraged recapitalization allow firms to gain a surplus from tax shields in a similar market equilibrium.

debt issuance leads to a reduction in equity issuance, no change in current payouts or cash on hand, and a reduction in future payouts due to additional interest expenses. Therefore, debt issuance, in this case, does save personal tax for shareholders, and Miller's formula applies in general. [13]

Since positive NPV investment opportunities are limited, there is a second case in which the firm has no positive NPV project to invest in but considers a capital restructure to issue debt and take advantage of the tax shield.[14] In this case, the firm must distribute the debt proceeds as payouts to reduce taxable income. Debt issuance transfers future payouts to current payouts and does not save personal taxes for shareholders as long as the expected future tax rates on payouts do not exceed the current tax rates. In this case, the tax benefits on each dollar of interest expense become the difference between the corporate tax rate and the personal tax rate on bond income, scaled by one minus the rate of unavoidable tax on equity income, $(1-\tau_e)(\tau_c-\tau_b)$. Miller's formula, $(1-\tau_b)-(1-\tau_c)(1-\tau_e)$, overstates the tax benefits from debt issuance in this scenario.

The differences are potentially significant and may flip the sign of tax benefits (or costs) from debt issuance. For example, if we apply the current top federal rates in 2022 (21% corporate tax, 37% personal tax on bond income, and 23.8% personal tax on equity income) to the definitions above, tax benefits per dollar of interest expense are (1 - 37%) - (1 - 21%)(1 - 23.8%) = 2.8% for financing investments and (21% - 37%)(1 - 23.8%) = -12.19% for capital restructuring. The use of 2017 rates before the tax reform (39% corporate tax, 40.79% personal tax on bond income, and 24.99% personal tax on

[13] Tax benefits can be lower than Miller's formula when the project's NPV is negative if financed by equity and when the firm expects to distribute no payouts in the near future.

[14] Securities such as treasury bonds are supposed to have the same risk-adjusted returns as the firm's bond and a zero NPV.

equity income) results in tax benefits of 13.45% vs. -1.34% for the two cases. A typical firm, therefore, would gain tax benefits from financing investments but not from capital structuring.

1.4. The model

The following model elucidates firms' optimal capital structure decisions when considering corporate and personal taxes and bankruptcy costs. The firm adjusts leverage freely without commitment as in DeMarzo and He (2021), at no cost of security issuance/repurchase. The key difference between this model and DeMarzo and He (2021) is that this model features personal taxes on bond and equity income. The leverage dynamics also differ from DeMarzo and He (2021) in several ways. First, there are two local leverage targets for firms with leverage above and below a threshold; the lower target is 0 when the corporate tax rate τ_c is lower than the personal tax rate on bond income τ_b. Second, if $\tau_c < \tau_b$, the firm repurchases debt when it has existing debt and more earnings than expenses or expects external financing in a relatively long future. And third, firms gain from tax benefits even without a commitment device.

1.4.1. Model setup

Agents are risk neutral with a discount rate r, and they face a corporate tax at rate τ_c, a personal tax on bond income at rate τ_b, and a personal tax on equity income at rate τ_e. A firm's EBIT follows:

$$dY_t = (\mu(Y_t) + I(Y_t))\,dt + \sigma(Y_t)dZ_t \tag{1.1}$$

Investment $I(Y_t)$ can be financed by internal cash flow and issuance of debt and equity. For simplicity, I assume investment is exogenous in the baseline model with a linear cost $\kappa I(Y_t)$. I study endogenous investment choice and debt overhang in Section 6.

The firm issues debt with a coupon rate $c > 0$ that matures exponentially at a rate $m > 0$. Let F_t be the face value of existing debt.[15] Then, the payment flow to debtholders is $(c+m)F_t dt$. Let Φ_t be the endogenous cumulative debt issuance, and $d\Phi_t < 0$ represents a debt repurchase. Existing debt evolves by $dF_t = d\Phi_t - mF_t dt$. Denote the price of debt issuance as $p(Y_t, F_t)$. Debtholders receive an after-tax cash flow of $[(1-\tau_b)c + m]F_t dt$ until bankruptcy. Assume the recovery value is zero at bankruptcy.

Let Γ_t be the endogenous cumulative proceeds from equity issuance, with $\Gamma_t \geq 0$ by definition.[16] To ensure a reasonable payouts policy that the firm does not save forever, assume that the firm does not hold cash.[17] The dividend flow can be written as:

(1.2)

$$d\Delta_t = [Y_t - \underbrace{\tau_c(Y_t - cF_t)}_{\text{corporate tax}} - \underbrace{(c+m)F_t}_{\text{payments to debt}} - \underbrace{\kappa I(Y_t)}_{\text{investment}}]dt + \underbrace{p(Y_t, F_t)d\Phi_t}_{\text{proceeds from debt}} + \underbrace{d\Gamma_t}_{\text{proceeds from equity}}$$

15 No Ponzi assumption: $F_t < \bar{F}(Y_t)$, where $\bar{F}(Y_t)$ exceeds the unleveraged value of the firm.

16 I assume there are no frictions in the equity market, so there is no need to characterize equity issuance prices. Maximizing value for all shareholders is equivalent to maximizing value for existing shareholders. As equity issuance and payouts are taxed differently, they are modeled separately here; payouts are taxed, but debt repurchase is not.

17 A firm does not hold cash if, for example, the firm earns no return on internal cash. Although internal equity and external equity are taxed differently in this model, such an assumption is not as harmful as it seems, as firms do not save cash to avoid equity issuance even if they are allowed to do so. To confirm this argument, assume that securities are priced by two state variables, EBIT and net debt outstanding. If the firm earns the same expected return on cash holdings as on its own debt, it cannot generate a higher cash flow from cash than from repurchasing debt. Buying back its own debt, which only defaults when the firm itself defaults, is optimal. Therefore, only zero-leverage firms, which need no external financing in the future, would save in this model if we relax the no cash assumption.

where $d\Delta_t \geq 0$ since dividends are nonnegative. The firm maximizes the expected net payouts to shareholders

$$V_t = \max_{T,\Phi,\Gamma} E\left[\int_t^T e^{-r(s-t)}\left[(1-\tau_e)d\Delta_s - d\Gamma_s\right]\right] \tag{1.3}$$

by choosing optimal capital structure over time and bankruptcy time T.

1.4.2. Security valuations

Substitute (1.2) into (1.3), the value function can be written as

(1.4)

$$V(Y_t, F_t) = \max_{T,\Phi,\Gamma} E_t\left[\int_t^T e^{-r(s-t)}\Big((1-\tau_e)\left\{\left[Y_s - \tau_c(Y_s - cF_s) - (c+m)F_s - \kappa I(Y_s)\right]ds + p(Y_s, F_s)d\Phi_s\right\}ds - \tau_e d\Gamma_s\Big)\right]$$

Observe that if a firm issues a dollar of equity and distributes a dollar of dividend simultaneously, there is a net loss of τ_e due to the personal income tax paid on the dividend. We then have the following rule for equity issuance:

Proposition 1. ***(Optimal equity issuance)*** *The firm does not issue equity and distribute dividends at the same time. Optimal equity issuance is uniquely pinned down by optimal debt issuance.*

$$d\Gamma_t = max\left\{-\left[Y_t - \tau_c(Y_t - cF_t) - (c+m)F_t - \kappa I(Y_t)\right]dt - p(Y_t, F_t)d\Phi_t, 0\right\} \tag{1.5}$$

When the firm issues equity, the proceeds are "trapped" in the firm and cannot be taken back by the shareholders without paying a personal income tax. Therefore, a firm that maximizes the total payoff of all shareholders should never issue equity and distribute dividends simultaneously. Equity financing policies can be characterized by net payouts, with negative payouts representing an equity issuance.

Following DeMarzo and He (2021), I look for a smooth equilibrium with continuous issuance policy, where $d\Phi_t = \phi_t dt$ and $d\Gamma_t = \gamma_t dt$.

Proposition 2. ***(Smooth equilibrium)*** *$V(Y,F)$ is strictly decreasing in F when $p(Y,F) > 0$. If for any F and F', $V(Y,F) > V(Y,F') + \left(1 - \mathbf{1}_{\{F'-F>0\}}\tau_e\right) p(Y,F)(F'-F)$, then Φ_t and Γ_t are continuous. If $V(V,F)$ is differentiable to F, then $(1-\tau_e)p(Y,F) \leq -V_F(Y,F) \leq p(Y,F)$. A sufficient condition for a smooth equilibrium is that $V(Y,F)$ is strictly convex in F.*

Proof. This proposition is true because the firm always has the option to adjust debt from F to F'. If $F' < F$, adjusting debt to F' discretely is financed by equity issuance. Otherwise it leads to a dividend distribution with a personal tax, so

$$
\begin{aligned}
V(Y,F) &\geq V(Y,F') + (F'-F)p(Y,F) - \mathbf{1}_{\{F'-F>0\}}\tau_e(F'-F)p(Y,F) \\
&= V(Y,F) + \left(1 - \mathbf{1}_{\{F'-F>0\}}\tau_e\right) p(Y,F)(F'-F)
\end{aligned}
\tag{1.6}
$$

When the inequality is strict, there is no discrete adjustment of leverage. Then following Proposition 1, there is no discrete equity issuance. The issuance policies are continuous. If $V(V,F)$ is differentiable to F, inequality (1.6) can be reorganized as $(1-\tau_e)p(Y,F) \leq -V_F(Y,F) \leq p(Y,F)$.

If $V(V,F)$ is strictly convex in F, that is, $V(Y,F) > V(Y,F') - V_F(Y,F)(F'-F)$, when $F' > F$,

$$V(Y,F) > V(Y,F') - V_F(Y,F)(F'-F) \geq V(Y,F') + (1-\tau_e)p(Y,F)(F'-F) \tag{1.7}$$

so there is no discrete debt adjustment. When $F' < F$,

$$V(Y,F) > V(Y,F') - V_F(Y,F)(F'-F) \geq V(Y,F') + p(Y,F)(F'-F) \tag{1.8}$$

there is no discrete debt adjustment either. Therefore, $V(V,F)$ strictly convex in F is sufficient for a smooth equilibrium. □

Now I derive optimal conditions for an equilibrium. I look for an equilibrium where security values p, V are twice continuously differentiable, and $p_F(Y,F) < 0$, which means debt price decreases with the amount of outstanding debt given EBIT. Given Proposition 1, it is enough to find conditions for optimal debt policies since dividends and equity issuance are pinned down by the state variables and debt policies.

Let $\bar{\phi}(Y,F) = -\frac{1}{p(Y,F)}\left[Y - \tau_c(Y-cF) - (c+m)F - \kappa I(Y)\right]$, representing the debt issuance/repurchase such that there is no dividend distribution or equity issuance. Consider when $F > 0$ so that debt repurchase is not bounded by 0. Then the HJB equation for the value function (1.4) is

(1.9)

$$rV(Y,F) = \max\Bigg(\max_{\phi \geq \bar{\phi}} \Bigg\{ (1-\tau_e) \underbrace{\left[Y - \tau_c(Y - cF) - (c+m)\,F - \kappa I(Y) + p(Y,F)\phi\right]}_{\text{positive net payouts}} + \underbrace{(\phi - mF)V_F(Y,F)}_{\text{debt evolution}} + \underbrace{\left(\mu(Y) + I(Y)\right)V_Y(Y,F) + \frac{1}{2}\sigma(Y)^2 V_{YY}(Y,F)}_{\text{earnings evolution}} \Bigg\},$$
$$\max_{\phi < \bar{\phi}} \Bigg\{ \underbrace{\left[Y - \tau_c(Y - cF) - (c+m)\,F - \kappa I(Y) + p(Y,F)\phi\right]}_{\text{negative net payouts (equity issuance)}} + \underbrace{(\phi - mF)V_F(Y,F)}_{\text{debt evolution}} + \underbrace{\left(\mu(Y) + I(Y)\right)V_Y(Y,F) + \frac{1}{2}\sigma(Y)^2 V_{YY}(Y,F)}_{\text{earnings evolution}} \Bigg\} \Bigg)$$

When $\phi(Y,F) > \bar{\phi}(Y,F)$, the firm is distributing a positive amount of dividend, so a personal tax on equity income at rate τ_e is charged on the payouts. In contrast, when $\phi(Y,F) < \bar{\phi}(Y,F)$, net payouts are negative, representing the proceeds from equity issuance, so there is no tax charged. If $\phi(Y,F) = \bar{\phi}(Y,F)$, net payouts are zero, and the first and second parts of the maximization are identical.

A necessary condition for a ϕ to be optimal is that either the first order condition holds or a constraint is binding, so an optimal debt issuance policy must satisfy either

$$p(Y,F) = -\frac{V_F(Y,F)}{1-\tau_e} \quad \text{and} \quad \phi > \bar{\phi}(Y,F) \tag{1.10}$$

or

$$-V_F(Y,F) \leq p(Y,F) \leq -\frac{V_F(Y,F)}{1-\tau_e} \quad \text{and} \quad \phi = \bar{\phi}(Y,F) \tag{1.11}$$

or

$$p(Y,F) = -V_F(Y,F) \quad \text{and} \quad \phi < \bar{\phi}(Y,F) \tag{1.12}$$

Since both parts of the value function's HJB equation are linear in $\phi(Y,F)$, these conditions are also sufficient. When the condition (1.10) holds, the firm distributes dividends and is indifferent to issuing extra debt for distributing dividends. When the condition (1.11) holds, the firm issues no equity and pays no dividends, issuing exactly enough debt to finance expenses or repurchasing debt with all free cash flow. When the condition (1.12) holds, the firm issues equity and is indifferent between debt and equity financing.

1.4.3. Optimal debt policies

Next, I derive debt issuance/repurchase rules when one of the first order conditions holds, using HJB equations for the value function and the debt price. I then determine if these financing policies are feasible and consistent with the inequalities comparing ϕ to $\bar{\phi}$ corresponding to the first order conditions in each case.

The equilibrium debt price satisfies

$$p(Y_t,F_t) = E_t\left[\int_t^T e^{-(r+m)(s-t)}\left[(1-\tau_b)c + \frac{1}{m}\right] ds\right] \tag{1.13}$$

The HJB equation for debt price (1.13) is

$$\begin{aligned} rp(Y,F) =& (1-\tau_b)c + m(1-p(Y,F)) + (\phi - mF)p_F(Y,F) \\ &+ (\mu(Y) + I(Y))\, p_Y(Y,F) + \frac{1}{2}\sigma(Y)^2 p_{YY}(Y,F) \end{aligned} \tag{1.14}$$

When the firm distributes dividends, $p(Y,F) = -\frac{V_F(Y,F)}{1-\tau_e}$, substitute it into the first item of the maximization in (1.9) and take derivative to F, we get

$$\begin{aligned} -rp(Y,F) =& \tau_c c - (c+m) + p_F(Y,F)\phi + mp(Y,F) - (\phi - mF)\, p_F(Y,F) \\ & - (\mu(Y) + I(Y))\, p_Y(Y,F) - \frac{1}{2}\sigma(Y)^2 p_{YY}(Y,F) \end{aligned} \tag{1.15}$$

Add (1.15) to (1.14), then

$$0 = (\tau_c - \tau_b)c + p_F(Y,F)\phi \tag{1.16}$$

Hence when $p(Y,F) = -\frac{V_F(Y,F)}{1-\tau_e}$ holds within a continuous region of (Y,F), the debt policy is

$$\phi = -\frac{(\tau_c - \tau_b)c}{p_F(Y,F)} = \frac{(1-\tau_e)(\tau_c - \tau_b)c}{V_{FF}(Y,F)} \tag{1.17}$$

Since $p_F(Y,F) < 0$, ϕ has the same sign as $\tau_c - \tau_b$. The firm repurchases debt when distributing dividends if $\tau_c < \tau_b$ and issues debt otherwise. Here $p(Y,F) = -\frac{V_F(Y,F)}{1-\tau_e}$ is the first order condition when the firm distributes dividends, so this debt policy is feasible to be an equilibrium choice when $\frac{(1-\tau_e)(\tau_c-\tau_b)c}{V_{FF}(Y,F)} \geq \bar{\phi}(Y,F)$. If $\tau_c < \tau_b$, the firm has enough earnings to finance this debt repurchase with internal cash. If $\tau_c > \tau_b$, this debt issuance raises more funds than the firm's financing need, and the rest is distributed as dividends.

Similarly, when the firm issues equity, $p(Y,F) = -V_F(Y,F)$, substitute it into the second item of the maximization in (1.9), take the derivative to F and add it to (1.14), we get

$$\phi = -\frac{(\tau_c - \tau_b)c}{p_F(Y,F)} = \frac{(\tau_c - \tau_b)c}{V_{FF}(Y,F)} \tag{1.18}$$

In this case, ϕ also has the same sign as $\tau_c - \tau_b$. The firm repurchases debt when issuing equity if $\tau_c < \tau_b$, and it issues debt otherwise. It is feasible if $\frac{(1-\tau_e)(\tau_c-\tau_b)c}{V_{FF}(Y,F)} \leq \bar{\phi}(Y,F)$, which means the firm repurchases more debt than earnings can fund or issues less debt than its financing need. Then, I summarize the firm's equilibrium debt policies as the following result.

Proposition 3. ***(Debt strategies conditional on security values)*** *Depending on the comparison between the debt price $p(Y,F)$ and the value function's marginal change to outstanding debt $V_F(Y,F)$, the firm's equilibrium debt policy $\phi(Y,F)$ is given by one of followings:*

1. If $p(Y,F) = -\frac{V_F(Y,F)}{1-\tau_e}$ within a continuous neighborhood of the state variables (Y,F),

$$\phi(Y,F) = \frac{(1-\tau_e)(\tau_c - \tau_b)c}{V_{FF}(Y,F)}$$

2. If $-V_F(Y,F) < p(Y,F) < -\frac{V_F(Y,F)}{1-\tau_e}$, or there exists $(Y+\epsilon_Y, F+\epsilon_F)$ arbitrarily close to (Y,F) that the inequalities hold,

$$\phi(Y,F) = \bar{\phi}(Y,F) = -\frac{1}{p(Y,F)}\left[Y - \tau_c(Y-cF) - \left(c + \frac{1}{m}\right)F - \kappa I(Y)\right]$$

3. If $p(Y,F) = -V_F(Y,F)$ within a continuous neighborhood of the state variables (Y,F),

$$\phi(Y,F) = \frac{(\tau_c - \tau_b)c}{V_{FF}(Y,F)}$$

In the first and third cases, when the firm distributes dividends or issues equity, debt policy $\phi(Y, F)$ can be represented as the marginal tax benefit of debt divided by the value function's second-order derivative to outstanding debt, as in DeMarzo and He (2021). However, the marginal tax benefit here, either $(1 - \tau_e)(\tau_c - \tau_b)$ or $(\tau_c - \tau_b)$ per a dollar of coupon payment, is different from both DeMarzo and He (2021) (τ_c) and the traditional definition with personal taxes $(1 - \tau_b) - (1 - \tau_c)(1 - \tau_e)$. In the first case, proceeds from additional debt issuance are distributed as dividends, so the personal tax on equity τ_e cannot be saved. In the third case, the firm expects to issue equity continuously and pay no dividends in the near future, when most coupons on additional debt are paid, so personal tax on equity income at rate τ_e is not paid or saved. The marginal tax benefits in both cases have the same sign as $(\tau_c - \tau_b)$, which is negative for top statutory rates in most years in the U.S. In contrast, the traditional definition of marginal tax benefits—with or without personal taxes—always leads to a positive benefit based on statutory tax rates.

If $V(Y, F)$ is strictly convex in F, $V_{FF}(Y, F) > 0$ and $\phi(Y, F)$ has the same sign as $\tau_c - \tau_b$ in these two cases, the firm repurchases debt when $\tau_c < \tau_b$. Although the tax benefits align with reducing bankruptcy risk when $\tau_c < \tau_b$, the firm does not repurchase as much as debt possible. This conclusion follows because the debt ratchet effect means that debt repurchase reduces the risk of existing debt and benefits the debtholders at the cost of shareholders. Such cost increases with debt repurchase, and the firm repurchases debt until it is indifferent with additional debt within a continuous neighborhood of (Y, F). With the equilibrium levels of debt repurchase, the firm is indifferent between using a marginal dollar of internal cash for dividends and debt repurchase in the first case

$((1-\tau_e)p(Y,F) = -V_F(Y,F))$, and indifferent between issuing a marginal dollar of debt and equity in the third case $(p(Y,F) = -V_F(Y,F))$.

In the second case when $-V_F(Y,F) < p(Y,F) < -\frac{V_F(Y,F)}{1-\tau_e}$, debt financing is cheaper than external equity but more expensive than internal cash. The firm prefers debt financing to equity financing, and it prefers debt repurchase to dividend distribution. It breaks even by issuing or repurchasing debt without any cash flow to or from the shareholders.

Proposition 3 shows the firms' debt policies conditional on the relations between the debt price $p(Y,F)$ and the value function's marginal change on debt $V_F(Y,F)$. We can further characterize debt strategies based on the state variables (Y,F) by checking the conditions at extreme values of parameters and the boundaries between regions of state variables where each of the above debt policies applies. Then I find general relations between the state variables (Y,F) and the comparison of $p(Y,F)$ and $V_F(Y,F)$ by continuity and monotonicity without further specifications of functional forms and solving $p(Y,F)$ and $V(Y,F)$. Proposition 4 summarizes the results, with proof in the appendix.

Proposition 4. ***(Optimal financing policies)*** *If $(1-\tau_c)Y - \kappa I(Y)$ monotonically increases in Y, the firm's optimal debt policies on the space of state variables (Y,F) can be divided into at most four continuous regions, with leverage from high to low:*

Region 1 (Equity issuing region): $p(Y,F) = -V_F(Y,F)$, the firm issues equity and issues/repurchases debt if $\tau_c - \tau_b$ is positive/negative.

Region 2 and 4 (Break even by debt regions): $-V_F(Y,F) < p(Y,F) < -\frac{V_F(Y,F)}{1-\tau_e}$, the firm issues no equity and distributes no dividends, and it breaks even by issuing/repurchasing debt if earnings are less/more than expenses.

Region 3 (Dividend distributing region): $p(Y,F) = -\frac{V_F(Y,F)}{1-\tau_e}$*, the firm issues equity and issues/repurchases debt if* $\tau_c - \tau_b$ *is positive/negative.*

Regions 2 always exists. Regions 3 must exist if $\tau_c \geq \tau_b$*.*

The figures below illustrate the results of Proposition 4 indicating the variation in the debt policies with the scaled interest coverage ratio Y/F for any given value of Y, in the cases when $\tau_c < \tau_b$ and $\tau_c \geq \tau_b$. The red line represents the bankruptcy-triggering levels of interest coverage ratio. On its right, when leverage is high and close to bankruptcy, the firm issues equity, and the first order condition $p(Y,F) = -\frac{V_F(Y,F)}{1-\tau_e}$ holds. The firm's debt policy is given by $\phi(Y,F) = \frac{(1-\tau_e)(\tau_c-\tau_b)c}{V_{FF}(Y,F)}$. When leverage is lower than in the region above, the firm prefers debt financing to equity financing and prefers debt repurchase to dividend distribution. $-V_F(Y,F) < p(Y,F) < -\frac{V_F(Y,F)}{1-\tau_e}$ and the debt policy is given by $\phi(Y,F) = \bar{\phi}(Y,F)$. The dotted purple line represents when the firm breaks even internally, with earnings exactly meeting expenses. When leverage is even lower, the firm distributes dividends and the first order condition $p(Y,F) = -V_F(Y,F)$ holds, and the debt policy is given by $\phi(Y,F) = \frac{(\tau_c-\tau_b)c}{V_{FF}(Y,F)}$. In the case that $\tau_c < \tau_b$, when leverage is low and close to zero, the firm spends all free cash flow on debt repurchase without distributing dividends until there is no outstanding debt.

In contrast to standard trade-off models in which the marginal source of financing is usually equity, the marginal source of financing is debt in the break-even by debt regions and equity in the other regions. This difference is the result of a wedge between the costs of internal and external equity due to taxes on dividends. When the cost of debt financing falls between the costs of internal and external equity, the firm has a pecking order preference prioritizing the use of internal equity to debt to external equity.

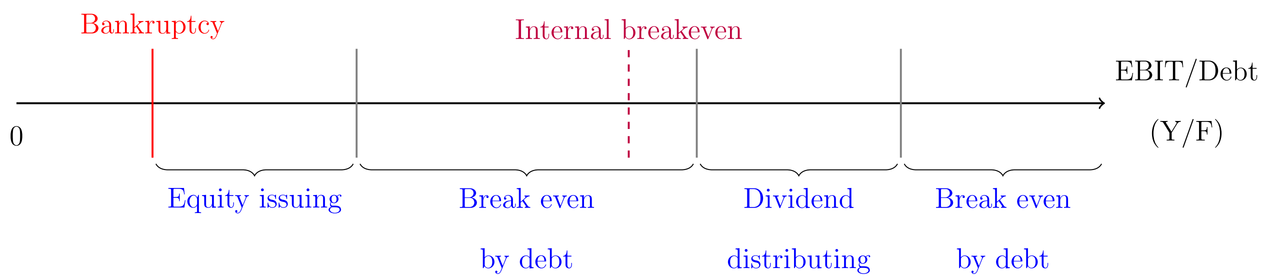


Figure: the firm's optimal financing policies, $\tau_c < \tau_b$

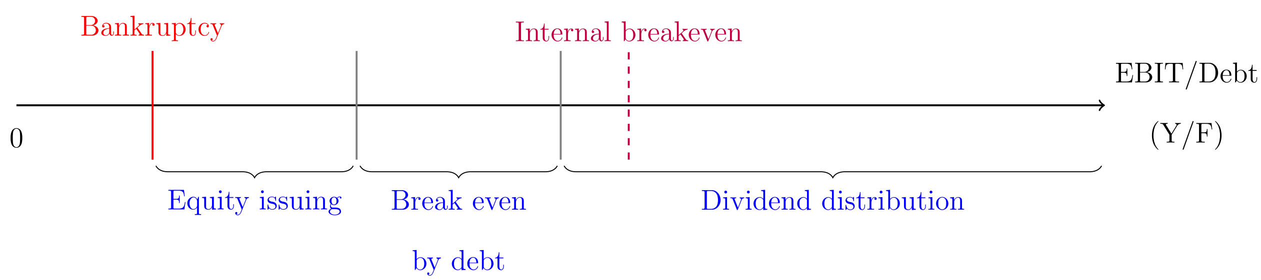


Figure: the firm's optimal financing policies, $\tau_c \geq \tau_b$

When $-V_F(Y, F) < p(Y, F) < -\frac{V_F(Y,F)}{1-\tau_e}$, the HJB equations of the value function are different from their counterparts with $\phi = 0$. Since $\tau_e > 0$ and the high leverage break-even by debt region always exists, the no-trade valuation in DeMarzo and He (2021) no longer holds in this model. Noticing that no trade of debt is always a feasible option for the firm, we have the following result.

Corollary 1. *As long as $\tau_e > 0$, the firm benefits from the tax shield of debt even if there is no commitment to future leverage policies, $V(Y, F) > V^0(Y, F)$ where $V^0(Y, F)$ represents the no-trade value of the firm.*

1.4.4. Tax benefits

When $\tau_c < \tau_b$, the firm repurchases debt not only when earnings exceed expenses but also when leverage is high and close to the bankruptcy threshold. This behavior occurs because coupons do not save personal equity income tax for shareholders if they are financed by equity issuance. To further understand tax benefits in this model, I discuss the expected value of total and marginal tax benefits below.

The value function can be decomposed as

$$V(Y,F) = V^{rf}(Y,0) + TB(Y,F) - BC(Y,F) \tag{1.19}$$

where the value of net corporate and personal income tax savings by debt follows

$$\begin{aligned} rTB(Y,F) =& (\tau_c - \tau_b)cF + \mathbf{1}_{\{\phi > \bar{\phi}\}}\tau_e \left[\left((1-\tau_c)c + m \right) F - p(Y,F)\phi \right] \\ &+ (\phi - mF)TB_F(Y,F) + (\mu(Y) + I(Y))TB_Y(Y,F) + \frac{1}{2}\sigma(Y)^2 TB_{YY}(Y,F) \end{aligned} \tag{1.20}$$

and the bankruptcy cost follows

$$rBC(Y,F) = (\phi - mF)BC_F(Y,F) + (\mu(Y) + I(Y))BC_Y(Y,F) + \frac{1}{2}\sigma(Y)^2 BC_{YY}(Y,F) \tag{1.21}$$

When the firm does not distribute dividends, the flow payoff is the difference between the corporate tax saved and the personal income tax paid on coupons. If $\tau_b > \tau_c$, such flow payoff is negative, representing a tax cost to the firm. When the firm distributes dividends, besides the cost of $\tau_c - \tau_b$ discounted by the personal income tax on equity τ_e, there is a saving of personal income tax on equity on the payments to debtholders net of proceeds from debtholders. The firm gains a tax benefit from existing debt but not from issuing new debt. When issuing new debt with a face value of a dollar, the firm's expected marginal tax benefit is

$MTB(Y, F) =$

(1.22)

$$-\mathbf{1}_{\{\phi_0 > \bar{\phi}_0\}} \tau_e p(Y, F) + E\left[\int_0^T e^{-(r+m)t}\left[(\tau_c - \tau_b)c + \mathbf{1}_{\{\phi_t > \bar{\phi}_t\}} \tau_e\left((1-\tau_c)c + m\right)\right] dt \middle| Y, F\right]$$

where $\phi(Y_t, F_t), \bar{\phi}(Y_t, F_t)$ are written as $\phi_t, \bar{\phi}_t$ for short. The marginal tax benefit is higher when the firm is not distributing dividends and when it is more likely to distribute dividends in the near future when coupons and principals are paid to debtholders. In an extreme case, if the firm is not distributing now but is expected to distribute at all time in the future, the marginal tax benefit on each dollar of coupon coin coincides with Miller's formula $(1 - \tau_b) - (1 - \tau_e)(1 - \tau_c)$. Miller's formula, then, represents an upper bound for the marginal tax benefits of interest payments after adjusting for personal taxes.

1.4.5. A roadmap for solving the general model

An equilibrium of the general model can be found by:

1. Find the bankruptcy threshold $\{Y, F\}_b$ and the boundaries between the equity issuing region and the break-even by debt region $\{Y, F\}_e$. Start with arbitrary positive initial values with higher leverage than the internal break-even level, and generate value and price functions by (1.9) and (1.14), setting

- $V^b = V^b_Y = V^b_F = 0$, $p^b = 0$,

- $\phi = \frac{(\tau_b - \tau_c)}{p_F(Y,F)}$ in the equity issuing region,
- $\phi = \bar{\phi} = -\frac{1}{p(Y,F)}\left[Y - \tau_c(Y - cF) - (c+m)F - \kappa I(Y)\right]$ if $p(Y,F) < -\frac{V_F(Y,F)}{1-\tau_e}$, and $\phi = \frac{(\tau_b - \tau_c)}{p_F(Y,F)}$ if $p(Y,F) = -\frac{V_F(Y,F)}{1-\tau_e}$ in the regions with lower leverage than the equity issuing region.

and check if the firm's value and debt price converge to their limits when Y goes to infinity for each F. Adjust the boundaries until reaching convergence.

2. Generate the equilibrium debt issuance ϕ, value function $V(Y,F)$ and debt price $p(Y,F)$ as above using the boundary conditions found.

3. Verify that $V(Y,F)$ and $p(Y,F)$ are strictly decreasing in F, $V(Y,F)$ is strictly convex in F, and $-V_F(Y,F) \leq p(Y,F) \leq -\frac{V_F(Y,F)}{1-\tau_e}$.

4. Generate equity issuance γ by (1.5).

1.5. Leverage dynamics

Next, I discuss the leverage dynamics in a homogenous case, such that all variables can be expressed as a function of a single state variable $y_t = \frac{Y_t}{F_t}$. I focus on the baseline scenario in which $1 - \tau_c \geq 1 - \tau_b > (1-\tau_c)(1-\tau_e)$.[18]

1.5.1. A homogenous model

Consider the case that $\mu(Y_t) = \mu Y_t, I(Y_t) = iY_t, \sigma(Y_t) = \sigma Y_t$, with $\mu + i < r$, $\tau_c + \kappa i < 1$. Define $y_t \equiv \frac{Y_t}{F_t}, v(y_t) \equiv V\left(\frac{Y_t}{F_t}, 1\right), p(y_t) = p\left(\frac{Y_t}{F_t}, 1\right)$. Then the model is homogenous in F_t

[18] I made $1 - \tau_c \geq 1 - \tau_b > (1-\tau_c)(1-\tau_e)$ the baseline assumption because such relation holds for the U.S. top federal statutory rates in most years.

and

$$V(Y_t, F_t) = V\left(\frac{Y_t}{F_t}, 1\right) F_t = v(y_t)F_t \tag{1.23}$$

$$p(Y_t, F_t) = p\left(\frac{Y_t}{F_t}, 1\right) = p(y_t) \tag{1.24}$$

Since

$$dY_t = iY_t dt + \sigma Y_t dZ_t \tag{1.25}$$

$$dF_t = (\phi_t - mF_t)dt \tag{1.26}$$

y_t follows

$$\frac{dy_t}{y_t} = \left(i + m - \frac{\phi_t}{F_t}\right) dt + \sigma dZ_t \tag{1.27}$$

We can rewrite the HJB equations (1.9) and (1.14) in y_t using

$$V_F(Y, F) = v(y) - yv'(y), \quad V_Y(Y, F) = v'(y), \quad V_{YY} = \frac{1}{F}v''(y) \tag{1.28}$$

$$p_F(Y, F) = -\frac{y}{F}p'(y), \quad p_Y(Y, F) = \frac{1}{F}p'(y), \quad p_{YY} = \frac{1}{F^2}p''(y) \tag{1.29}$$

Then the HJB equation for the value function (1.9) becomes

$$rv(y) = \max\Bigg(\max_{\phi\geq\bar{\phi}} \Bigg\{ (1-\tau_e)\left[y - \tau_c(y-c) - (c+m) - \kappa i y + p(y)\phi\right] + \left(\frac{\phi}{F} - m\right)\left[v(y) - yv'(y)\right] + iyv'(y) + \frac{1}{2}\sigma^2 y^2 v''(y) \Bigg\},$$
$$\max_{\phi<\bar{\phi}} \Bigg\{ \left[y - \tau_c(y-c) - (c+m) - \kappa i y + p(y)\phi\right] + \left(\frac{\phi}{F} - m\right)\left[v(y) - yv'(y)\right] + iyv'(y) + \frac{1}{2}\sigma^2 y^2 v''(y) \Bigg\} \Bigg) \tag{1.30}$$

and the HJB equation for the price function (1.14) becomes

$$rp(y) = (1-\tau_b)c + m(1-p(y)) - (\phi - mF)\frac{y}{F}p'(y) + \lambda i y p'(y) + \frac{1}{2}\sigma^2 y^2 p''(y) \tag{1.31}$$

I solve the model by finding the boundaries between regions of financing strategies and the bankruptcy threshold, such that the value and price functions converge to their limits as y_t goes to infinity, following the roadmap in the previous section. Denote y_b as the bankruptcy threshold and $y_e \geq y_b$ as the boundary between the equity issuing region and the high-leverage break-even by debt region. In the equity issuing region, when $y < y_e$, $p(y) = -v(y) + yv'(y)$, $\phi = \frac{(\tau_c - \tau_b)cF}{yp'(y)}$ and the firm issues equity, the HJB equations are

$$(r+m)\,v(y) = (1-\tau_c-\kappa i)y - (1-\tau_c)c - m + (m+i)\,yv'(y) + \frac{1}{2}\sigma^2 y^2 v''(y) \tag{1.32}$$

$$(r+m)\,p(y) = (1-\tau_c)c + m + (m+i)\,yp'(y) + \frac{1}{2}\sigma^2 y^2 p''(y) \tag{1.33}$$

In the break-even by debt regions, when $y \geq y_e$, $\frac{1}{1-\tau_e}[-v(y)+yv'(y)] \geq p(y) \geq -v(y)+yv'(y)$, $\phi = -\frac{F}{p(y)}\left[(1-\tau_c-\kappa i)y-(1-\tau_c)c-m\right]$, the HJB equations are

(1.34)
$$\begin{aligned}(r+m)\,v(y) =& -\frac{1}{p(y)}\left[(1-\tau_c-\kappa i)y-(1-\tau_c)c-m\right]\left[v(y)-yv'(y)\right]+(m+i)\,yv'(y)\\&+\frac{1}{2}\sigma^2y^2v''(y)\end{aligned}$$

(1.35)
$$\begin{aligned}(r+m)\,p(y) =&(1-\tau_b)c+m+\left[(1-\tau_c-\kappa i)y-(1-\tau_c)c-\frac{1}{m}\right]y\frac{p'(y)}{p(y)}+(m+i)\,yp'(y)\\&+\frac{1}{2}\sigma^2y^2p''(y)\end{aligned}$$

In the dividend distributing region, when $y \geq y_e$ and $p(y) = \frac{1}{1-\tau_e}[-v(y)+yv'(y)]$, $\phi = \frac{(\tau_c-\tau_b)cF}{yp'(y)}$ and the firm distributes dividends, the HJB equations are

(1.36)
$$(r+m)\,v(y) = (1-\tau_e)[(1-\tau_c-\kappa i)y-(1-\tau_c)c-m]+(m+i)\,yv'(y)+\frac{1}{2}\sigma^2y^2v''(y)$$

(1.37)
$$(r+m)\,p(y) = (1-\tau_c)c+m+(m+i)\,yp'(y)+\frac{1}{2}\sigma^2y^2p''(y)$$

A solution of the model can be pinned down by the following boundary conditions

$$\lim_{y\to\infty} p(y) = \frac{(1-\tau_b)c+m}{r+m} \qquad \lim_{y\to\infty} v(y) = \frac{(1-\tau_e)(1-\tau_c-\kappa i)y}{r-i} \tag{1.38}$$

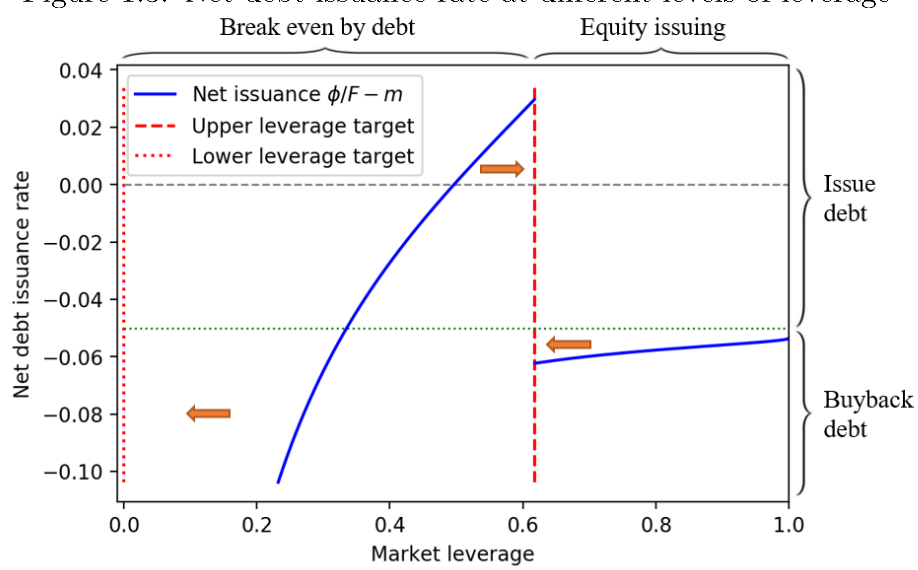
$$p(y_b) = 0 \qquad v(y_b) = 0 \tag{1.39}$$

and smooth pasting conditions that $v'(y_b) = 0$, $v'(y)$ and $p'(y)$ are continuous at the boundaries between regions. The differential equations (1.34) and (1.35) have no known closed-form solutions. Therefore, I solve $v(y)$ and $p(y)$ numerically by picking y_b, y_e and generating function values from y_b to infinity using the HJB equations and the bankruptcy values, then check if the values converge to their closed-form limits above. The functions are generated by a fourth-order Runge-Kutta-Nystrm algorithm, with details in the appendix.

1.5.2. Optimal debt policies and leverage targets

Figure 1.3 shows the firm's optimal net debt issuance rate $\frac{\phi}{F} - m$ as a function of its market leverage, defined by $\frac{\text{book value of debt}}{\text{market value of equity + book value of debt}} = \frac{F}{V(Y,F)+F} = \frac{1}{v(y)+1}$, in a baseline case with $r = 5\%, \tau_c = 30\%, \tau_b = 35\%, \tau_e = 20\%, \mu = 0, i = 2\%, \sigma = 40\%, c = \frac{r}{1-\tau_b}, m = 5\%, \kappa = 20$. Parameters of the geometric Brownian motion $\mu + i = 2\%$ and $\sigma = 40\%$ follows DeMarzo and He (2021). $\kappa i = 40\%$ is chosen based on the aggregate ratio of net investments (capital expenditures net of depreciation) to EBIT of Compustat firms. Negative net issuance rates represent debt repurchasing. Unlike the firm in DeMarzo and He (2021) which never repurchases debt, the firm repurchases debt both when leverage is high and when leverage is low. Since I am assuming that investments are linear in earnings in this homogenous case, the firm's net financing need $-[(1 - \tau_c - \kappa i)y - (1 - \tau_c)c - m]$ is monotonically increasing in leverage. When financing need is non-positive, the firm cannot save personal tax on equity for shareholders by debt since the proceeds are also taxed when distributed to shareholders as dividends. When financing need is high, the firm does not expect to pay dividends and the taxes on dividends in the near future. Therefore, in both scenarios, debt issuance depends on the comparison between the corporate and personal tax rates $\tau_c - \tau_b$, and the firm repurchases debt if $\tau_c < \tau_b$. The firm only issues debt in the high-leverage break-even by debt region when it has a moderate level of leverage. Debt issuance jumps up at the boundary between the equity issuing region and the break-even by debt region when the firm switches from equity financing to debt financing.

Figure 1.3. Net debt issuance rate at different levels of leverage



Notes. The figure plots the firm's net debt issuance rate $\phi/F - m$, given its current market leverage $1/(v(y)+1)$. The figure shows two local leverage targets at 0.62 and 0. The firm's leverage converges to 0.62 when it is above 0.5 and 0 otherwise. The parameters are $r = 5\%, \tau_c = 30\%, \tau_b = 35\%, \tau_e = 20\%, \mu = 0, i = 2\%, \sigma = 40\%, c = \frac{r}{1-\tau_b}, m = 5\%, \kappa = 20$.

Unlike the traditional wisdom of the trade-off theory that firms should have a single leverage target, this model allows for two local leverage targets. When leverage is high, the firm's leverage converges to a "upper leverage target"—the boundary between the equity issuing region and the high-leverage break-even by debt region—by issuing debt when leverage is below target and repurchasing debt when leverage is above target. When leverage is low, the firm converges to zero leverage as a "lower leverage target" by repurchasing debt until it reaches zero. Proposition 5 shows the conditions for the model to have two leverage targets.

Proposition 5. ***(Leverage targets)*** *If (1)* $\tau_c < \tau_b$ *or (2)* $m > 0$, $\tau_c = \tau_b$*, then 0 is a local leverage target. If* $y_e < \frac{(1-\tau_c)c+m(1-p(y_e))}{1-\tau_c-\kappa i}$*, the model has two local leverage targets, and a market leverage of* $\frac{1}{v(y_e)+1}$ *is also a local leverage target.*

1.5.3. Leverage dynamics

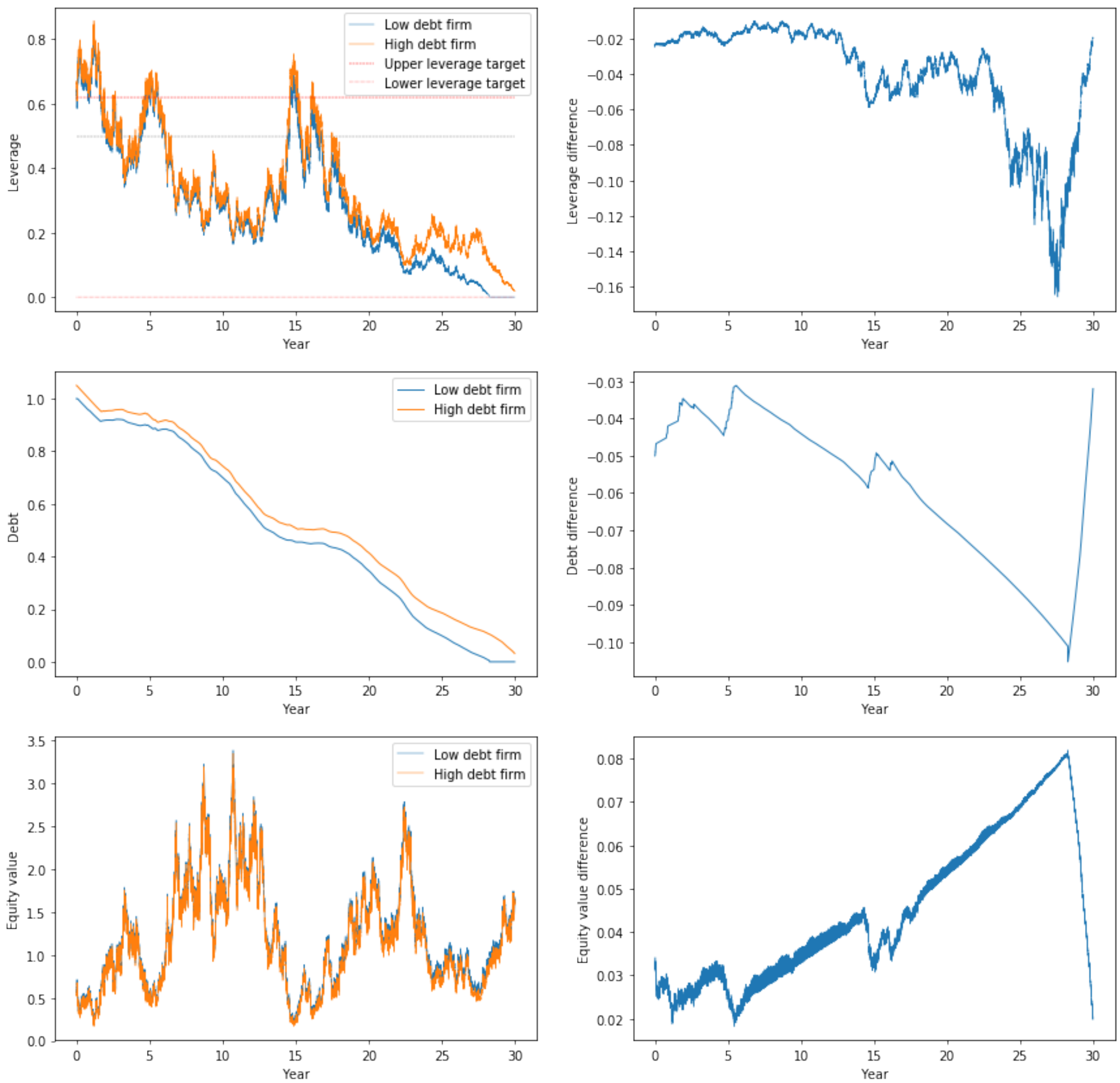
Since marginal tax benefits depend on the firm's financing needs, the firm's financing strategy depends on earnings and existing debt. The asymmetry in savings on personal equity income tax from debt can be an another force—in addition to the debt ratchet effect—that makes debt policies past-dependent. For example, when a firm is in the break-even by debt region, it borrows as much as its financing needs, including payments to debt outstanding.

Figure 1.4 illustrates an example of the models simulated leverage dynamics with the baseline parameter values. To highlight the potential persistence of the leverage differences between firms, I show the evolutions of two firms' leverage where both firms have the same earnings process Y_t all-time but a slightly different initial debt level. One has 5% more initial debt than the other, which may arise due to an earnings shock. For example, suppose both firms operate at the upper leverage target before time 0, and one has slightly higher earnings than the other. In that case, a negative earnings shock to the firm with higher earnings can make their earnings equal while leaving them with different levels of outstanding debt. Figures on the left show the evolutions of the two firms' market leverage $\left(\frac{1}{v(y)+1}\right)$, face value of outstanding debt (F) and equity values $(v(y)F)$, and figures on the right show the differences between the two firms.

Both firms start with leverage close to the upper leverage target. Their leverages deviate from the target after receiving earnings shocks and adjust to targets slowly. While both firms start at the upper leverage target, they converge to the lower target at the end of the 30-year period, and the low-debt firm reaches zero leverage. Importantly, their leverages can cross the boundaries between regions so that their leverage target can switch between upper and lower targets.

Although the two firms have the same earnings flow, their leverage and valuation differences persist throughout the 30-year period and even grow larger. Before the low-debt firm reached zero leverage, the leverage difference reached 16%, and the equity value difference reached 8%—more than double the initial differences that were below 4%. A temporary shock, then, can have long-persisting effects on a firm's capital structure. Such persistence implies that cross-sectional differences in leverage between firms can continue for a long time period, even without differences in earnings.

Figure 1.4. Leverage dynamics and persistence of differences



Notes. The figure plots the simulated leverage dynamics in 30 years of two firms with identical EBIT flows but starting with different levels of debt. Leverage adjusts to targets slowly. A firm's leverage target can switch from one to the other due to earnings shocks, until reaching zero leverage. Differences in leverage, debt and equity value can persist for a long time. The parameters are $r = 5\%, \tau_c = 30\%, \tau_b = 35\%, \tau_e = 20\%, \mu = 0, i = 2\%, \sigma = 40\%, c = \frac{r}{1-\tau_b}, m = 5\%, \kappa = 20$.

1.5.4. A Simulation of leverage distribution

To further explore the model's implications for the cross-sectional differences in firms' leverage, I simulate the cross-sectional distribution of firms' leverage in the model and compare it to the data. To generate a stationary leverage distribution and allow zero-leverage firms to fail, I assume that firms face an exogenous random Poisson shock such that EBIT drops to zero. Such a shock can be interpreted as, for example, the firm's product becoming outdated due to competitors technological advance.[19] Each bankruptcy firm, either due to debt payments or the technology shock, is replaced by a new firm that makes a lump-sum investment to enter and chooses the optimal fractions of debt and equity to finance the investment.

Let λ be the density of the Poisson technology shock dN_t. Then we can write the earnings process as

$$dY_t = (\mu + i)Y_t dt + \sigma Y_t dZ_t - Y_t^- dN_t \tag{1.40}$$

where Y_t^- denotes earnings before the shock. The HJB equations become

[19]In reality, earnings are not always positive as in a geometric Brownian motion. I match the rate of such shocks to the proportion of Compustat firms with earnings that drop from positive to negative and remain negative for at least three consecutive years (about 2%).

$$rv(y) = \max\left(\max_{\phi\geq\bar{\phi}}\left\{(1-\tau_e)\left[y-\tau_c(y-c)-(c+m)-\kappa iy+p(y)\phi\right]\right.\right.$$
$$\left.+\left(\frac{\phi}{F}-m\right)\left[v(y)-yv'(y)\right]+iyv'(y)+\frac{1}{2}\sigma^2y^2v''(y)-\lambda v(y)\right\},$$
$$\max_{\phi<\bar{\phi}}\left\{\left[y-\tau_c(y-c)-(c+m)-\kappa iy+p(y)\phi\right]\right.$$
$$\left.\left.+\left(\frac{\phi}{F}-m\right)\left[v(y)-yv'(y)\right]+iyv'(y)+\frac{1}{2}\sigma^2y^2v''(y)-\lambda v(y)\right\}\right) \tag{1.41}$$

For a new firm that needs to invest I_0 to enter and receive an earnings flow starting with Y_0, it chooses debt and equity financing to maximize the continuation value of the firm plus the payoff to shareholders at entry, that is,

$$v\left(y_0\right)F_0+\left(1-\mathbf{1}_{p(y_0)F_0-I_0>0}\tau_e\right)\left[p\left(y_0\right)F_0-I_0\right] \tag{1.42}$$

Since $y=\frac{Y}{F}$ is the only state variable in this homogenous case, I normalize initial earnings as $Y_0=1$. When I_0 is large enough, the firm gains a full tax shield without being constrained from saving personal tax on equity income. The following result describes the firm's optimal debt issuance.

Proposition 6. ***(Unconstrained optimal leverage)***
Let $F_0^*=\arg\max_{F_0}\left[v\left(\frac{1}{F_0}\right)+p\left(\frac{1}{F_0}\right)\right]F_0$ be the debt issuance that maximizes the firm's total enterprise value. If $p\left(\frac{1}{F_0^*}\right)F_0^*\leq I_0\leq\left[p\left(\frac{1}{F_0^*}\right)+v\left(\frac{1}{F_0^*}\right)\right]F_0^*$, the firm's optimal debt issuance is F_0^*.

Proof.

$$\begin{aligned}\left[v\left(\frac{1}{F_0^*}\right)+p\left(\frac{1}{F_0^*}\right)\right]F_0^*-I_0 &\geq \left[v\left(\frac{1}{F_0}\right)+p\left(\frac{1}{F_0}\right)\right]F_0-I_0 \\ &\geq \left[v\left(\frac{1}{F_0}\right)+p\left(\frac{1}{F_0}\right)\right]F_0-I_0-\mathbf{1}_{p(y_0)F_0-I_0>0}\tau_e\left[p\left(\frac{1}{F_0}\right)F_0-I_0\right]\end{aligned} \tag{1.43}$$

□

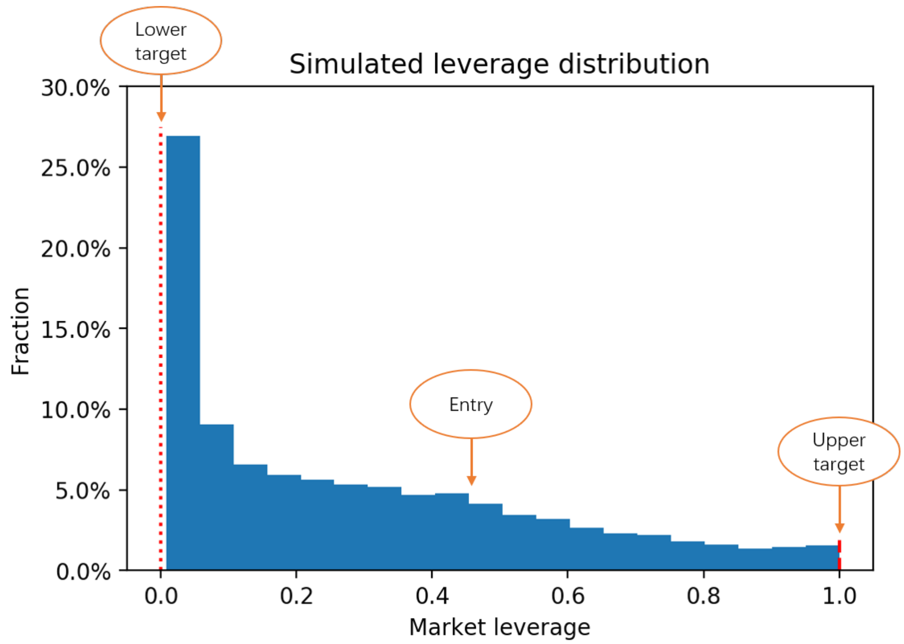
I assume that the initial investment for new firms satisfies the conditions above. Then, I simulate the leverage dynamics of 5,000 firms that start with the unconstrained optimal leverage. Suppose any firm fails due to the interest coverage ratio y falling below the bankruptcy threshold y_b or the technology shock. In that case, it is replaced by a new firm starting with the unconstrained optimal leverage. I simulate the evolutions of firms' leverage until it reaches a stationary distribution, with parameters $r = 5\%, \tau_c = 30\%, \tau_b = 30\%, \tau_e = 15\%, \mu = 0, i = 2\%, \sigma = 40\%, c = \frac{r}{1-\tau_b}, m = 1/15, \kappa = 20, \lambda = 2\%$. Here the Poisson shock density λ matches the proportion of Compustat firms with earnings that drop from positive to negative and remain negative for at least three consecutive years.

As exhibited in Figure 1.5, the simulated leverage distribution closely matches the leverage distribution of Compustat firms in the data in 1.1 when both are measured by Book value of debt/(Book value of debt + Market value of equity). As in the data, about 1/4 of firms have lower than 5% market leverage, and the fractions of firms in each 5% bin are decreasing in leverage levels.[20] Such a distribution contrasts with the bell-shaped

[20] A difference is that the simulated distribution has a thicker tail. The thinner tail in the data may be generated by more realistic assumptions about financially distressed firms, such as allowing for restructuring. I leave that for future work.

distribution implied by traditional trade-off models that have a single global leverage target.

Figure 1.5. Simulated leverage distribution



Notes. The figure plots a simulated stationary distribution of 5,000 firms' market leverage. Failed firms are replaced by new firms entering at the unconstrained optimal leverage. The distribution closely matches the market leverage distribution in the data shown in Figure 1.1. The parameters are $r = 5\%, \tau_c = 30\%, \tau_b = 30\%, \tau_e = 15\%, \mu = 0, i = 2\%, \sigma = 40\%, c = \frac{r}{1-\tau_b}, m = 1/15, \kappa = 20, \lambda = 2\%$.

1.5.5. Empirical implications

The leverage dynamics above have several empirical implications for firms' financing policies that align with existing empirical evidence or can be tested in the data. First, the model generates a reasonable fraction of zero-leverage firms and helps explain the zero-leverage puzzle (Strebulaev and Yang, 2013). The potentially zero or negative tax benefits of leveraged recapitalization rationalize zero-leverage firms' reluctance to increase leverage. Such an explanation can be tested by measuring firms' marginal tax benefits for issuing additional debt in the data following equation (1.22), which I study in Sections 7 and 8. The marginal benefits depend on firms' financing needs and can be negative with reasonable tax rates when there is no need for external financing. The traditional measure $(1-\tau_b)-(1-\tau_c)(1-\tau_e)$ overestimates marginal tax benefits for issuing additional debt, especially for firms without external financing need.

Second, leverage dynamics in the model are past-dependent with persistent differences, which is consistent with empirical evidence of persistent effect from past capital structure decisions (e.g., Baker and Wurgler, 2002) and persistent cross-sectional differences (e.g., Lemmon et al., 2008). Both the financing needs to pay debtholders and the debt ratchet effect make firms with higher leverage continue to issue more debt.

Third, in contrast to a traditional trade-off theory, the model features two local leverage targets instead of one. This difference allows the model to generate new empirical implications for firms' adjustments of their leverage to targets. A firm's leverage target can switch between two targets with significant differences due to changes in its leverage. Firms actively adjust leverage to targets when leverage is high (equity issuing region) regardless of financing needs, but they only passively adjust leverage with a financing pecking order of $internal\ cash \prec debt\ financing \prec equity\ financing$ when leverage is low. Such behavior largely differs from the traditional understanding of the trade-off theory that a firm always actively—although perhaps slowly or infrequently—adjusts to a single leverage target.

1.6. Welfare analysis and optimal taxation

In the traditional definition, interest expenses generate tax savings at a constant rate $(1-\tau_b)-(1-\tau_c)(1-\tau_e)$. In that case, taxing shareholders at the firm and personal levels are equivalent, given a fixed value of $(1-\tau_c)(1-\tau_e)$. However, this paper shows that corporate and personal taxes on equity income affect firms' financing policies differently. As a result, analyzing the effect of tax rates τ_c and τ_e on economic efficiencies in this model can lead to important policy implications. In addition, I discuss optimal maturity from the firm and social welfare perspectives.

1.6.1. Tax efficiency

In order to analyze the distortion of taxes on firms' capital structure and the resulting inefficient bankruptcy loss, I first decompose the firm's value into the values of equity, debt, expected tax revenue, and expected bankruptcy loss. Normalizing $F = 1$, the unleveraged pre-tax value of the firm equals $v_0^{pre-tax}(y) = y(1-\kappa I)/(r-I)$. Equity and debt values are $v(y)$ and $p(y)$. Expected bankruptcy loss is

$$BL(y) = BC(y)\frac{1-\kappa I}{(1-\tau_e)(1-\tau_c-\kappa I)} \tag{1.44}$$

where $BC(y)$ solves HJB equation (1.21). And expected tax revenue equals

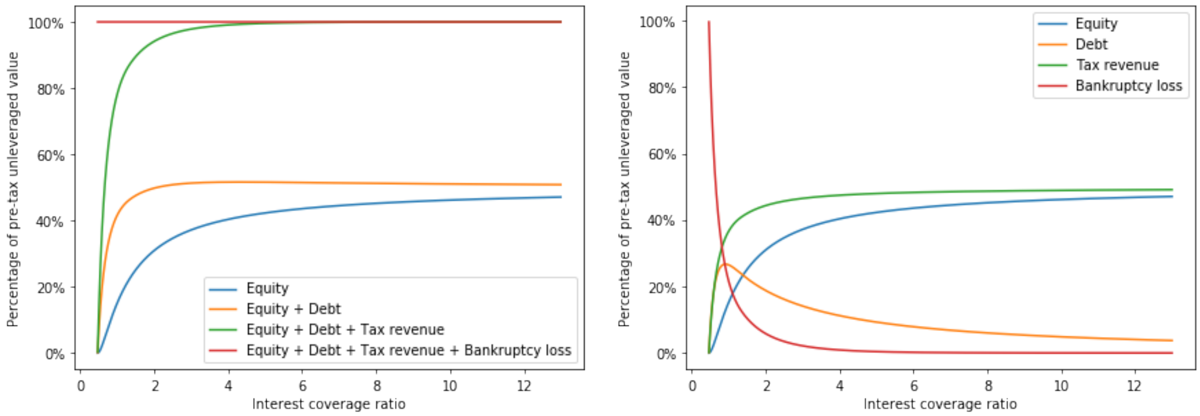
$$TR(y) = v_0^{pre-tax}(y) - v(y) - p(y) - BL(y) \tag{1.45}$$

Figure 1.6 shows the decompositions of the firm's pre- and after-tax values at different interest coverage ratios (y/c) with parameters $r = 5\%, \tau_c = 30\%, \tau_b = 35\%, \tau_e = 20\%, \mu =$

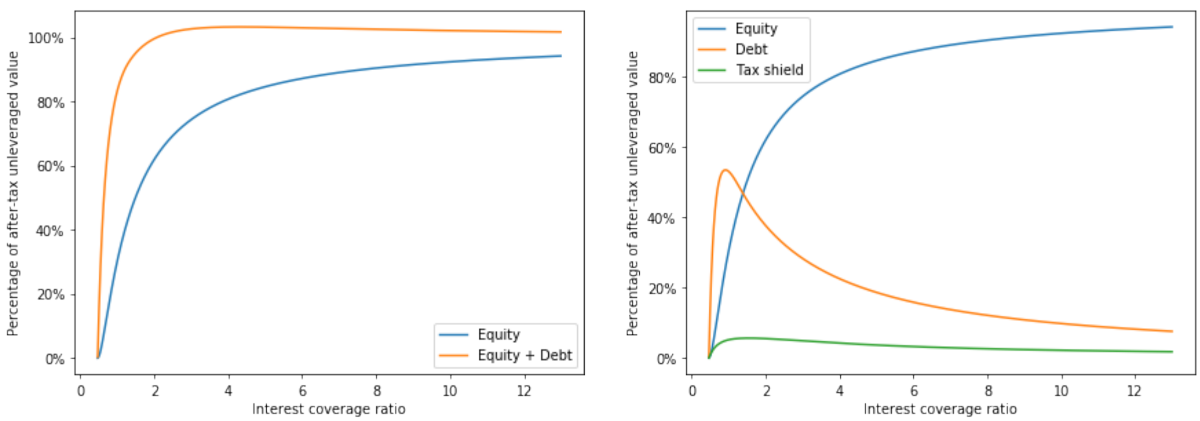
$0, i = 2\%, \sigma = 40\%, c = \frac{r}{1-\tau_b}, m = 5\%, \kappa = 10$. As the interest coverage ratio increases, there is lower bankruptcy risk, higher expected tax revenue, and a shift of the firm's capital structure from debt to equity. Taxes take up to 60% of the firm's earnings net of investments. The value of debt peaks at 26.73% when the interest coverage ratio is 0.92. Total enterprise value (equity + debt) peaks at the unconstrained optimal leverage when the interest coverage ratio is 4.31. In that case, the firm earns 3.26% of its after-tax unleveraged value from the tax shield of debt net of expected bankruptcy cost. The tax shield of debt is worth 5.67% of the firm's pre-tax unleveraged value and 5.58% of the firm's after-tax unleveraged value at the maximum.

Figure 1.6. Decomposition of the firm's value

(a) Pre-tax value decomposition



(b) After-tax value decomposition

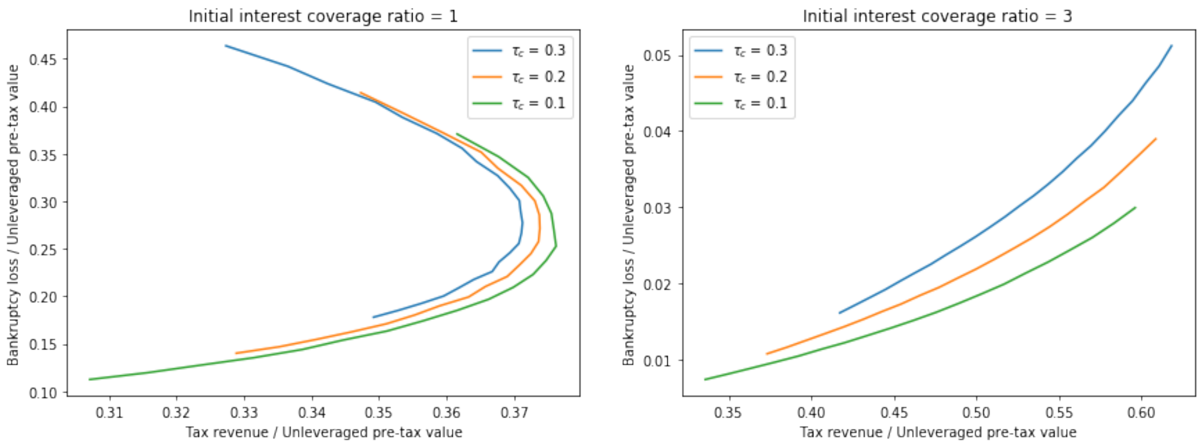


Notes. This figure shows decompositions of the firm's value at different interest coverage ratios (y/c) before and after taxes. Panel (a) decomposes the firm's pre-tax unleveraged value into the values of equity, debt, expected tax revenue, and expected bankruptcy loss. Panel (b) plots the firm's equity and debt value normalized by its after-tax unleveraged value. Pictures on the left plot the cumulative sum of the components, and pictures on the right plot values of each component. The parameters are $r = 5\%, \tau_c = 30\%, \tau_b = 35\%, \tau_e = 20\%, \mu = 0, i = 2\%, \sigma = 40\%, c = \frac{r}{1-\tau_b}, m = 5\%, \kappa = 10$.

A government may want to collect more tax revenue or a target level of tax revenue while minimizing the deadweight bankruptcy loss caused by leverage distortions. The following analysis focuses on the case in which the policymaker chooses a combination of corporate tax rate τ_c and personal tax rate on equity income τ_e, given a fixed personal tax rate on bond income τ_b. If the government can choose all three tax rates τ_c, τ_b, τ_e freely, it can set τ_b high enough that firms never use debt. The government can then tax an arbitrary proportion of the firm's earnings without causing inefficiency. In the real world, however, personal tax rates on bond income are usually the same as the rates on wages. It is reasonable to take such personal tax rates as given in the problem discussed here since these rates involve other redistributive concerns that are not covered in this paper.

Figure 1.7 plots feasible pairs of expected bankruptcy loss and tax revenue, normalized by the firm's unleveraged pre-tax value for different combinations of corporate tax rate τ_c and personal tax rate on equity income τ_e. Assume that the personal tax rate on bond income is fixed at $\tau_b = 35\%$. Each line plots the feasible sets with different values of τ_e when the corporate tax rate τ_c is 10%, 20%, and 30%. Pictures on the left and right plot the cases in which the firm's interest coverage ratio is 1 and 3, as examples for high and low leverage firms. The other parameters are $r = 5\%, \mu = 0, i = 2\%, \sigma = 40\%, c = \frac{r}{1-\tau_b}, m = 5\%, \kappa = 10$. The government aims to achieve outcomes in the lower right with higher tax revenue and lower expected bankruptcy loss. The upward-sloping parts of the lines represent the desirable choices where the government faces a trade-off between tax revenue and expected bankruptcy loss. The figure shows that for both high- and low-leverage firms, a lower corporate tax rate can push the line of feasible outcomes to the right, which is preferred.

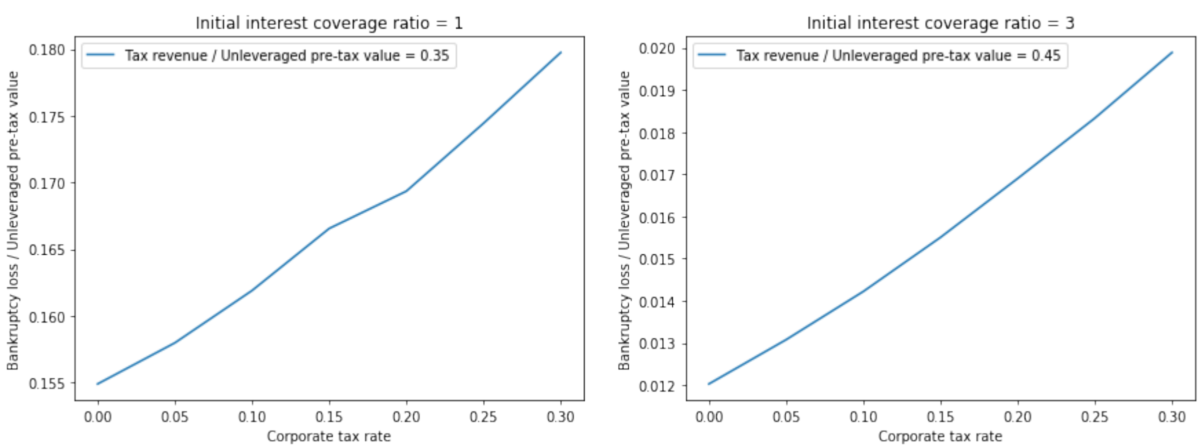
Figure 1.7. Feasible pairs of expected bankruptcy loss and tax revenue



Notes. This figure shows feasible pairs of expected bankruptcy loss and tax revenue, both normalized by the firm's unleveraged pre-tax value, for different combinations of corporate tax rate τ_c and personal tax rate on equity income τ_e. Personal tax rate on bond income is fixed at $\tau_b = 35\%$. Each line plots the feasible sets with different values of τ_e given $\tau_c = 10\%/20\%/30\%$. Pictures on the left and right plot the cases when the firm's interest coverage ratio is 1 and 3, as examples for high- and low-leverage firms. The other parameters are $r = 5\%, \mu = 0, i = 2\%, \sigma = 40\%, c = \frac{r}{1-\tau_b}, m = 5\%, \kappa = 10$.

In Figure 1.8, I study the optimal choice of corporate tax rate when the policymaker has a fixed tax revenue target. The figure plots the expected bankruptcy loss normalized by the firm's unleveraged pre-tax value when the policymaker collects a target tax revenue with different corporate tax rates τ_c. In this case, the personal tax rate on equity income τ_e is automatically pinned down by the target tax revenue and the other tax rates. The picture on the left plots the case in which the firm's interest coverage ratio is one, and the target tax revenue is 35% of the firm's unleveraged pre-tax value. This case serves as an example of a high-leverage firm. The picture on the right plots the case in which the firm's interest coverage ratio is three and the target tax revenue is 45% of the firm's unleveraged pre-tax value. This case is an example of a low-leverage firm. Other parameters are the same as above. Given the tax revenue targets, expected bankruptcy loss increases with the corporate tax rate in both cases. Therefore, the optimal approach to collecting tax revenue in these cases is to set the corporate tax rate at 0 and only tax the shareholders with the personal tax on equity income. In general, the government can reduce expected bankruptcy loss due to leverage distortions without losing tax revenue by taxing shareholders more at the personal level and less at the corporate level.

Figure 1.8. Optimal corporate tax rate given target tax revenue

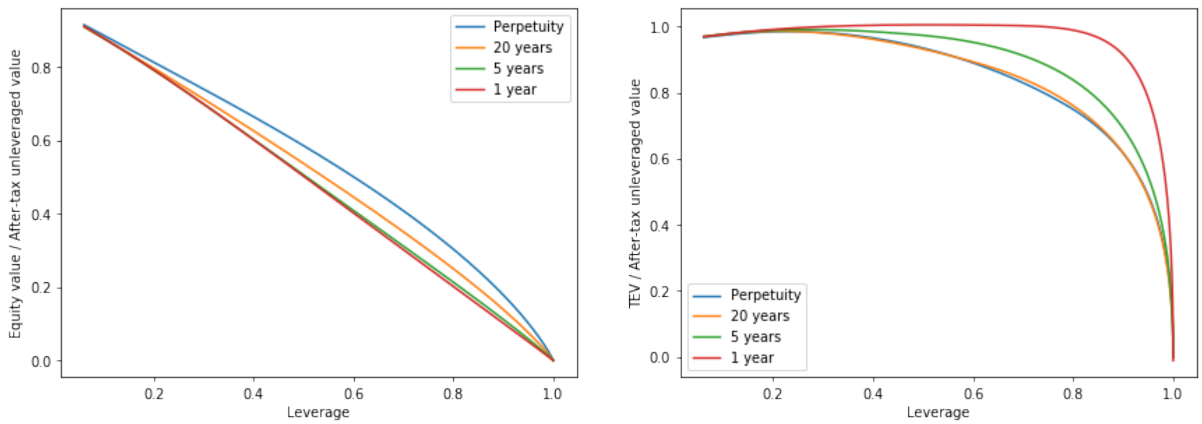


Notes. This figure shows the expected bankruptcy loss normalized by the firm's unleveraged pre-tax value when the policymaker collects a target tax revenue with different corporate tax rates τ_c. Personal tax rate on bond income is fixed at $\tau_b = 35\%$. Personal tax rate on equity income τ_e is automatically pinned down by the target tax revenue and the other tax rates. The picture on the left plots the case in which the firm's interest coverage ratio is 1 and the target tax revenue is 35% of the firm's unleveraged pre-tax value as an example of a high-leverage firm. The picture on the right plots the case in which the firm's interest coverage ratio is 3 and the target tax revenue is 45% of the firm's unleveraged pre-tax value as an example of a low-leverage firm. The other parameters are $r = 5\%, \mu = 0, i = 2\%, \sigma = 40\%, c = \frac{r}{1-\tau_b}, m = 5\%, \kappa = 10$.

1.6.2. Optimal maturity

Similarly, we can assess the firm's maturity preference by comparing the values of securities with different maturity rates. Figure 1.9 plots the firm's security values when it issues debt with different fixed maturity rates. The picture on the left plots equity values normalized by the firm's after-tax unleveraged value at different leverage levels when the expected maturity is 1, 5, 20 years, or infinity. The picture on the right plots the total enterprise value (debt + equity) normalized by the firm's after-tax unleveraged value at different leverage levels when the expected maturity is 1 year, 5 years, 20 years, or infinity. The equity value increases with expected maturity, but total enterprise value decreases with expected maturity because longer maturity lowers rollover risk for firms and transfers risk from shareholders to debtholders. Without other frictions, firms issue long-term debt, which is suboptimal from a social welfare perspective and leads to higher leverage and bankruptcy risk.

Figure 1.9. Security values with different debt maturity rates



Notes. This figure shows the firm's security values when it issues debt with different fixed rates of maturity. The picture on the left plots equity values normalized by the firm's after-tax unleveraged value at different levels of leverage, when the expected maturity is 1 year, 5 years, 20 years, or infinity. The firm's equity value is higher when debt maturity is longer, given any level of leverage. The picture on the right plots total enterprise value (debt + equity) normalized by the firm's after-tax unleveraged value at different levels of leverage, when the expected maturity is 1 year, 5 years, 20 years, or infinity. The firm's total enterprise value is higher when debt maturity is shorter, given any level of leverage.

1.6.3. Endogenous investment and debt overhang

Here, I assume that investments are endogenous with quadratic costs $\frac{1}{2}\kappa i^2 Y$, and I study optimal investment rates and debt overhang. In this case, we can rewrite the HJB equation as

$$
\begin{aligned}
rv(y) = \max \Bigg(& \max_{\phi \geq \bar{\phi}} \Bigg\{ (1-\tau_e) \left[y - \tau_c(y-c) - (c+m) - \frac{1}{2}\kappa i^2 y + p(y)\phi \right] \\
& \qquad + \left(\frac{\phi}{F} - m \right) [v(y) - yv'(y)] + iyv'(y) + \frac{1}{2}\sigma^2 y^2 v''(y) \Bigg\}, \\
& \max_{\phi < \bar{\phi}} \Bigg\{ \left[y - \tau_c(y-c) - (c+m) - \frac{1}{2}\kappa i^2 y + p(y)\phi \right] \\
& \qquad + \left(\frac{\phi}{F} - m \right) [v(y) - yv'(y)] + iyv'(y) + \frac{1}{2}\sigma^2 y^2 v''(y) \Bigg\} \Bigg)
\end{aligned}
\tag{1.46}
$$

Taking the first order condition to i, the optimal investment is

$$
i = \frac{v'(y)}{\left(1 - 1_{\phi > \bar{\phi}} \tau_e\right) \kappa} \tag{1.47}
$$

In addition to the standard result, the optimal investment rate is discounted by $1_{\phi > \bar{\phi}} \tau_e$. The firm invests more to avoid taxes on distribution.

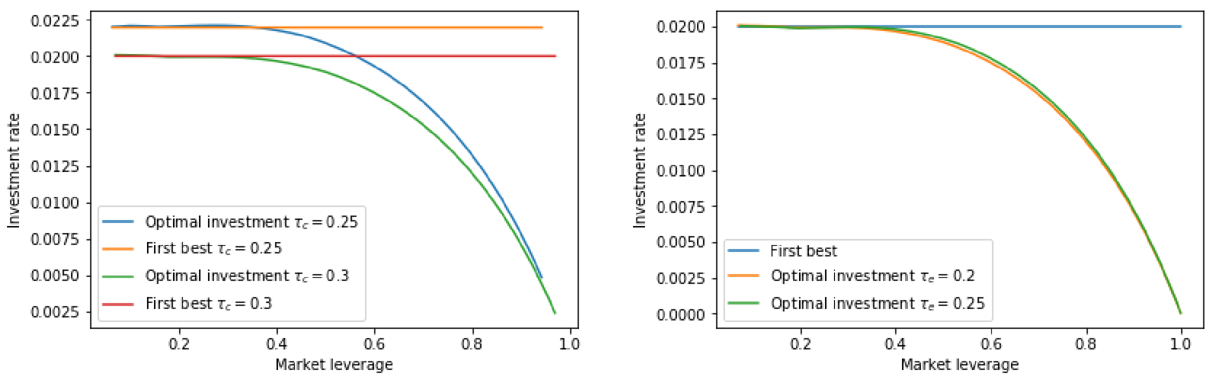
When the firm is unleveraged, the optimal investment rate is

$$
i_{unlev} = r - \mu - \sqrt{(r-\mu)^2 - \frac{2(1-\tau_c)}{\kappa}} \tag{1.48}
$$

Figure 1.10 plots firms' optimal investment rates at different levels of leverage, compared to the first best investment without leverage, and compared across different tax rates. Panel (a) plots the investment rates when the corporate tax rate is 25% and 30%. Panel (b) plots the investment rates when the personal tax rate on equity income is 20% and 25%. Investment rates decrease in leverage due to debt overhang. Given the level of leverage, a corporate tax cut improves investment, while a cut on the payout tax slightly reduces investments. This conclusion can be interpreted as the short-term effect of tax cuts since leverages adjust slowly and the indirect effect of tax cuts through their impact on leverages happens in the long term.

Figure 1.10. Equilibrium investment rates with differnt tax rates

(a) Investment with different corporate tax rates (b) Investment with different dividend tax rates



Notes. This figure shows the equilibrium investment rates compared to the first best investment rates at different levels of leverage. Panel (a) plots the investment rates when the corporate tax rate is 25% and 30%. Panel (b) plots the investment rates when the personal tax rate on equity income is 20% and 25%. Investment rates decrease in leverage due to debt overhang. A higher corporate tax rate leads to lower equilibrium and first best investment rates. A higher personal tax rate on equity income leads to slightly higher investment rates.

1.7. Conclusion

This paper shows that the tax benefits of debt for financing investment and for leveraged recapitalization are different. A dynamic trade-off model with corporate and personal taxes and bankruptcy costs features two local leverage targets. Depending on a firm's leverage compared to a threshold, it adjusts to one target or the other and may switch targets due to earnings shocks. When the corporate tax rate is lower than the personal tax rate on bond income, the lower leverage target is 0, which helps explain the zero-leverage puzzle (Strebulaev and Yang, 2013) that over 1/5 of firms have close to zero leverage. A simulation of the model generates a cross-sectional leverage distribution that closely matches the data.

The paper also studies policymakers choice of tax rates with a trade-off between tax revenue and expected deadweight bankruptcy loss due to leverage distortions. I show that policymakers can reduce expected bankruptcy loss without losing tax revenue by taxing shareholders more at a personal level and less at the corporate level. In the short run, a corporate tax cut increases investments, but a personal tax cut on equity income decreases investments.

CHAPTER 2

Benefits or costs? Measuring the tax consequences of debt from a dynamic perspective

2.1. Introduction

The traditional wisdom of debt tax shield suggestions that tax-paying firms can always gain tax benefits from shielding their earnings by interest expenses. However, previous literature (e.g. Graham, 2000, Strebulaev and Yang, 2013) find profitable firms with good liquidity seem underleveraged. A potential explanation is that the conventional definition of corporate tax savings from debt net of personal taxes following Miller (1977) does not apply to the real-world dynamic problem in which debt issuance may lead to additional payouts instead of a reduction in equity issuance (Hennessy and Whited, 2005, Hu, 2023).

Following Hu (2023), I build a novel empirical measure of marginal tax benefits from debt issuance and estimate it for Compustat firms during the period between 1980 and 2021 at the firm-year level. The measure assesses the marginal tax consequences of issuing new debt on current and future cash flows. I find that a substantial proportion of firms that seem underleveraged according to the traditional measure would see no benefit to levering up.

The measure differs from Miller's formula estimated by Graham (2000) in two ways. First, it takes personal taxes into account in a more realistic way because firms cannot save personal tax on income from equity when proceeds from debt would be distributed as payouts. Second, it measures the ex-ante marginal tax benefit from issuing new debt that increases future interest expenses, while Graham (2000) estimates the ex-post marginal tax rate by adding interest expense to realized current-year earnings. I also estimate the ex-post marginal tax rates for comparison. I measure marginal corporate tax rates following Graham (2000) and Blouin, Core, and Guay (2010) to account for detailed tax codes such as the carryforward and carryback of net operating losses.

The measure generates lower marginal tax benefits and a smaller mass of underleveraged firms than previous measures that use Miller's formula. Based on the new measure, 9.85% of firms face a positive marginal tax benefit, and 4.67% of firms can double debt before tax benefits decline. By contrast, the traditional measure indicates that 35.35% of firms face a positive marginal tax benefit and 20.36% of firms can double debt before tax benefits decline. The large mass of zero-leverage firms no longer seems a mystery—only about 5% of firms with below 5% book leverage can gain positive tax benefits from raising leverage. The new measure also explains the observed leverage changes in the data better than the traditional measure.

This paper is most closely related to the empirical literature on tax benefits and the underleveraged puzzle. Graham (2000) estimates the marginal tax rates for adding interest expenses to firms' earnings, following Miller's definition of tax benefits of debt. He notes that many firms seem underleveraged, and he finds that a typical firm can double its debt without leading to a reduction in the marginal tax benefit of interest expenses. Blouin, Core, and Guay (2010) show that a better simulation method for firms' future earnings partly resolves the puzzle and generates a smaller mass of underleveraged firms.[1] Strebulaev and Yang (2013) document that about 1/5 of firms have close to zero leverage that cannot be explained by the previous literature. They all find that underleveraged firms tend to have high profits, cash balances, and dividend distributions, which is perplexing from the perspective of financial distress costs. This paper's new marginal tax benefit measure incorporates the equity income tax costs on proceeds from debt issuance, showing that a typical payouts-distributing firm faces no tax benefits from a leveraged recapitalization.

This paper also contributes to the literature about the relation between tax rates and firms' leverage. Graham (1996a) finds a positive relation between firms' incremental use of debt and estimations of marginal corporate tax rates. Heider and Ljungqvist (2015) exploit staggerred changes in U.S. state tax rates and find that firms' leverage are sensitive to tax rate changes. Firms raise leverage in response to increases in state corproate tax rates. Faulkender and Smith (2016) study the relation between leverage and taxes at multinational firms and find firms operating in countries with higher corporate tax rates take higher leverage. I provide new empirical evidence that supports the postive relation between firms' leverage and tax consequences of debt issuance. Taking personal tax consequences of debt issuance into account, in additional to corporate taxes, makes the relation between leverage and marginal tax benefits even stronger.

The remainder of the paper is organized as follows. Section 2 builds an empirical measure of marginal tax benefits. Section 3 discusses the estimation results. Section 4

[1] Estimating firms' marginal corporate tax rates requires simulations of future earnings to capture the effect of carryback and carryforward rules for operating losses. Shevlin (1990) and Graham (1996aa, 1996bb, 2000) assume that future earnings follow random walks. Blouin, Core, and Guay (2010) show that random walks lead to biased estimations and propose a non-parametric simulation method.

shows that the new measure can better explain the observed leverage changes than the traditional measure. Section 5 concludes.

2.2. An empirical measure of marginal tax benefits

In this section, I build a new empirical measure of marginal tax benefits for a firm to issue an additional dollar of debt. The measure captures the different tax consequences of issuing debt to finance investment and recapitalizing the firm. As a result, I can estimate the extent to which we can explain the data, especially the low-leverage puzzle, by simply revising our view of tax benefits.

2.2.1. Measuring marginal tax benefits of debt issuance

The previous literature (e.g., Graham, 2000) measures the marginal tax benefits of debt by estimating firms' marginal corporate tax rates net of personal tax penalty, that is

$$(1 - \tau_b) - (1 - \tau_c)(1 - \tau_e) \tag{2.1}$$

This measure differs from the actual marginal tax benefits firms can earn from issuing new debt in two major ways. First, it does not capture the personal tax consequence of distributing the proceeds from debt to shareholders. In order to generate additional interest expense net of interest income, a firm either reduces equity issuance or distributes additional payouts when issuing debt. There is an increase in shareholders' personal tax on equity income if the firm's new debt issuance is expected to finance payouts.

Second, firms cannot add interest expenses to reduce current realized taxable income. If a firm issues new debt at the end of the year after observing the realized earnings, the new debt increases its interest expenses in the coming years until it matures.

To include these features in my estimation, following the theoretical work in Hu (2023), I define the marginal tax benefits measure as:

$$E^0 \left[\sum_{t=1}^{M} \frac{1}{(1+(1-\tau_b)r)^t} r \left[(1-\tau_b)(1-\mathbf{1}_{pay}\tau_{e,i0}) - (1-\tau_{c,it})(1-\tau_{e,it}) \right] \right] \tag{2.2}$$

Expression (2.2) defines the marginal tax benefits from issuing a debt that pays a dollar of coupon each period for M periods. Each dollar of interest expense generates proceeds from debt by the present value of $(1-\tau_b)$, where τ_b is the marginal debtholder's personal income tax rate that is priced in bonds. It increases the firm's current net payouts to shareholders by the present value of $(1-\tau_b)(1-\mathbf{1}_{pay}\tau_{e,i0})$, where $\mathbf{1}_{pay}=1$ when proceeds from additional debt are expected to be distributed as payouts. On the other hand, each dollar of interest expense reduces the firms' net payouts by $(1-\tau_{c,it})(1-\tau_{e,it})$.

Assume that firms expect no changes in tax policies. Firms then take the marginal debtholder's personal tax rate τ_b as given. Because of the gradient tax rates and allowance for carryback and carryforward of net operating losses, $\tau_{c,it}$ depends on firms' specific earnings expectation. Since a dollar of net payouts is taxed differently when used for reducing equity issuance, repurchasing stock, and paying dividends, $\tau_{e,i0}, \tau_{e,it}$ depends on the expectations of the forms of net payouts. I discuss measurement of these tax rates below.

2.2.2. Marginal corporate tax rates

I measure the marginal corporate tax rates following Graham (2000) and Blouin, Core, and Guay (2010). Since net operating losses (NOL) can be carried back or forward to offset taxable income in other years, estimating the corporate tax rates requires not only current taxable income but also an estimate of expected earnings in the future.[2] For example, estimating time t marginal tax benefits for an observation before 1997 requires simulating the earnings from $t+1$ to $t+18$. Graham (2000) simulates future earnings by a random-walk process $\Delta EBIT = \mu + \epsilon$. Blouin, Core, and Guay (2010) simulates future earnings by a non-parametric method in which firms estimate future growth in ROA and average assets by the growth rates of other firms with similar ROA and average assets and show that this method matches earnings growth in the data better. I account for taxable incomes using Compustat data following Appendix A in Blouin, Core, and Guay (2010) and adjust taxable incomes for NOL carryback and carryforward following Appendix A in Graham (2000). In Appendix C, I report the summary statistics of my replications of pre-financing marginal corporate tax rates using the two simulation methods in Table (B.1). In the following analysis, I use the measure based on the non-parametric simulation of earnings growth as the baseline measure.

[2]During the sample periods in this paper (1980-2021), the tax rules allow for (1) a 3-year carryback and 15-year carryforward in 1980-1997; (2) a 2-year carryback and 20-year carryforward in 1998-2017; and (3) no carryback, but indefinitely carry forward for up to 80% of taxable income since 2018. In addition, the CARES Act provided for a special 5-year carryback.

2.2.3. Marginal personal tax rates on bond income

Which investors' personal tax rates should firms take into account? Below, I assume firms maximize shareholder value, and I discuss the personal tax consequences in the two scenarios with positive and negative net payouts.

When net payouts are positive, proceeds from debt issuance are expected to be distributed to shareholders. Without changing their current consumption, shareholders can rebalance their portfolio by investing the payouts into a fixed-income security that earns the same return as the firm's debt issued. In that case, such recapitalization has an arbitrage-like effect on the shareholders' after-tax incomes. After the payouts at issue, they earn $(1-\tau_b)(1-\tau_e)$ more from interest income and $(1-\tau_c)(1-\tau_e)$ less from equity income per each dollar of interest expense on the new debt. The marginal tax benefits for the firm to issue new debt depend on comparing shareholders' marginal personal tax rates and the firms' marginal corporate tax rates. This finding is consistent with findings in Strebulaev and Yang (2013) that firms owned by family or with a large CEO ownership share tend to be zero-leveraged. Wealthy owners usually face the highest tax brackets and cannot benefit from leveraged recapitalizations.

When net payouts are negative, proceeds from debt issuance are expected to replace equity issuance. This replacement reduces the dilution of share value for existing shareholders. In this case, it is the marginal debtholder's personal rate priced in bonds that affects the amount the firm can raise from debt issuance and should be used for measuring tax benefits.

Due to the difficulty of observing shareholders' personal income tax rates and marginal bondholders' rates priced in bonds, I use the top federal statutory rates as a proxy. In the municipal bonds literature, Ang, Bhansali, and Xing (2010), Longstaff (2011), and Kueng (2018) find the short-term implied tax rates in bonds are close to the top rates, and Schwert (2017) uses the top statutory rates as a proxy for implied tax rates. Since measures of implied tax rates based on market prices are usually too volatile over time and sometimes exceed the top rates, using the top federal rates as a proxy can generate a more time-consistent estimate of marginal tax benefits.[3] A caveat is that this proxy may be biased for estimating the tax consequences of leveraged recapitalizations. That is because tax-exempt shareholders can potentially gain the difference between the priced tax rates and their own rates from investing the payouts. A better estimate requires observing the composition of shareholders and their personal income tax rates.

2.2.4. Marginal personal tax rates on equity income

Each dollar of a firm's cash inflows is subject to a dividend tax when directly paid to shareholders as dividends or a capital gains tax (with potential deferral benefits) if retained or used for share repurchase. Therefore, $\tau_{e,i0}$ and $\tau_{e,it}$ are weighted averages of dividend and capital gains tax rates depending on the proportion of the payouts or earnings that are expected to be distributed as dividends.

$\tau_{e,i0}$ is the expected marginal tax rate for current payouts in year 0 when the firm issues debt. Assuming that the firm uses the same payouts strategy as in the past year, I estimate it by $\tau_{e,i0} = \frac{dividends_{i0}}{payouts_{i0}}\tau_{div,0} + \left(1 - \frac{dividends_{i0}}{payouts_{i0}}\right)\tau_{cg,0}\alpha$. Here, $dividends_{i0}$ and $payouts_{i0}$ are

[3]For example, Ang, Bhansali, and Xing (2010) estimate an implied tax rate of over 70% based on transactions in the secondary market. Longstaff (2011) estimates the implied tax rates to be above the top rates in 2007 and around 10% during the financial crisis. An exception is Graham's (2000) measure using data from Fortune (1996). However, Fortune (1996) uses data from Salomon Brothers that is no longer available.

observed dividends and payouts paid in year 0, $\tau_{div,0}$ and $\tau_{cg,0}$ are the dividends and capital gains tax rates in year 0, and $\alpha = 0.25$, following Graham (2000), captures the benefits of deferring capital gains.

$\tau_{e,it}$ is the expected marginal tax rate on the reduced earnings by future interest expenses. To be conservative, I assume the marginal earnings are used for payouts, that is, $\tau_{e,it} = \tau_{e,i0}$. Alternatively, we may assume that the marginal earnings reduced by interest expenses are not necessarily distributed, yielding lower estimates of marginal tax benefits.[4]

2.3. Estimated marginal tax benefits from debt issuance

This section discusses findings from estimating marginal tax benefits from debt issues at the firm-year level. Compared to the traditional measure, the new measure described above better rationalizes many firms decisions to keep leverage low, indicating that most firms have adequately exploited the potential tax shield from debt.

2.3.1. Data

I use annual Compustat data for firm fundamentals. The sample includes firms headquartered in the US with at least three years of observations in 1980-2021, excluding those in the financial industry and American depository receipts ("ADR"). There are 199,436 firm-year observations.

[4]For example, one alternative is to assume firms only distribute the marginal earnings when earnings are positive.

I use Moody's Baa bond yields as the discount rates. Corporation income tax brackets and rates are from the Internal Revenue Service website. Top personal income tax rates are from Daniel Feenberg's taxsim website.

2.3.2. Summary statistics

Table 2.1 reports the summary statistics for the estimated marginal tax benefits with the new measure (2.2) and the traditional measure (2.1) for firm-year observations with positive taxable income. The full sample's summary statistics are reported in B.2 in Appendix C. I estimate the measures in two ways: (1) the ex-ante tax benefits for firms to raise a dollar of new perpetual debt after observing the current year's earnings; and (2) the ex-post tax benefits for firms to add a dollar of interest expense directly to the observed current-year earnings. Pre-financing and post-financing measures estimate the marginal tax benefits before and after accounting for the firms' actual interest expenses. Measures using the new definition generate lower marginal tax benefits than those using the traditional definitions. The average pre-financing marginal tax benefits are -7.12% for these firms with the new measure, compared to 2.86% with the traditional measure. This difference occurs because firms typically face negative marginal tax benefits from leveraged recapitalization with the new measure. 9.85% firms face a positive marginal tax benefit to issuing more debt than they actually did, compared to 35.35% based on the traditional measure.

Table 2.1. Descriptive statistics for marginal tax benefits

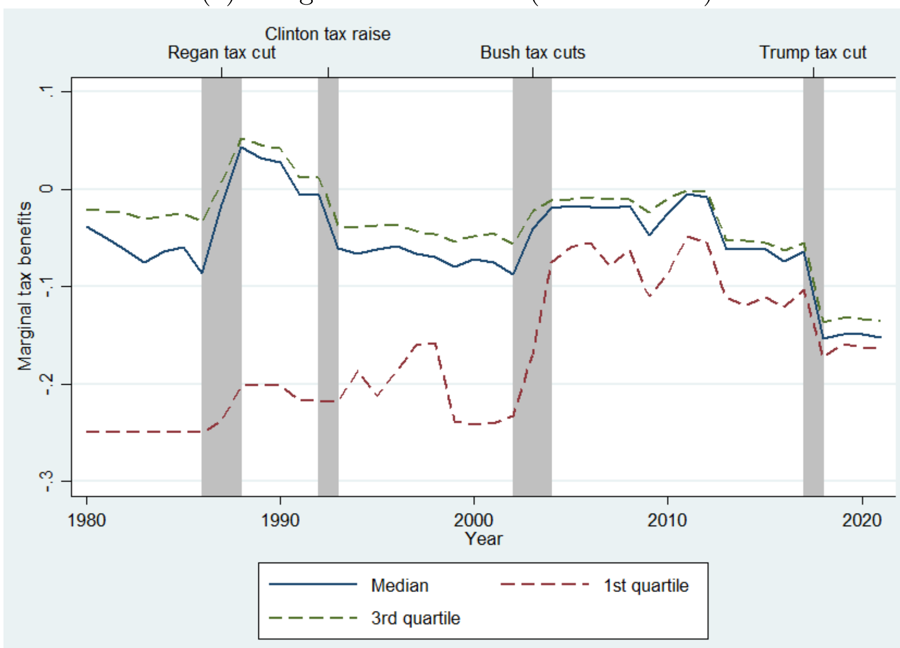
	Ex-ante prefin		Ex-post prefin		Ex-ante postfin		Ex-post postfin	
	new	traditional	new	traditional	new	traditional	new	traditional
obs	134,402	134,402	134,402	134,402	134,402	134,402	134,402	134,402
mean	-7.12%	1.25%	-5.5%	2.86%	-10.65%	-2.29%	-8.45%	-0.08%
std	11.79%	14.13%	11.67%	13.83%	13.93%	16.64%	13.79%	16.34%
min	-47.5%	-45%	-47.5%	-45%	-47.5%	-45%	-47.5%	-45%
1st quartile	-9.9%	-4.6%	-6.29%	-2.33%	-20.16%	-8.69%	-14.15%	-4.83%
median	-4.46%	0	-3.04%	1.47%	-5.83%	-0.62%	-3.84%	0
3rd quartile	-1.25%	8.78%	-0.3%	9.95%	-2.04%	7.33%	-0.77%	9.16%
max	25.26%	25.28%	33.12%	36.72%	25.26%	25.28%	33.12%	36.72%

Notes. This table shows the summary statistics for the marginal tax benefits for observations with positive taxable income in the 1980-2021 sample period. Ex-ante measures estimate the tax benefits for firms to raise a dollar of new perpetual debt after observing the current year earnings. Ex-post measures estimate the tax benefits for firms to add a dollar of interest expense directly to the observed current year earnings. Pre-financing and post-financing measures estimate the tax benefits before and after accounting for the firms' actual interest expenses, respectively. New definitions are those taking the personal tax consequence of leveraged recapitalization into account, as discussed above. Traditional definitions are given by $(1-\tau_b)-(1-\tau_c)(1-\tau_e)$. See appendix for summary statistics for the full sample.

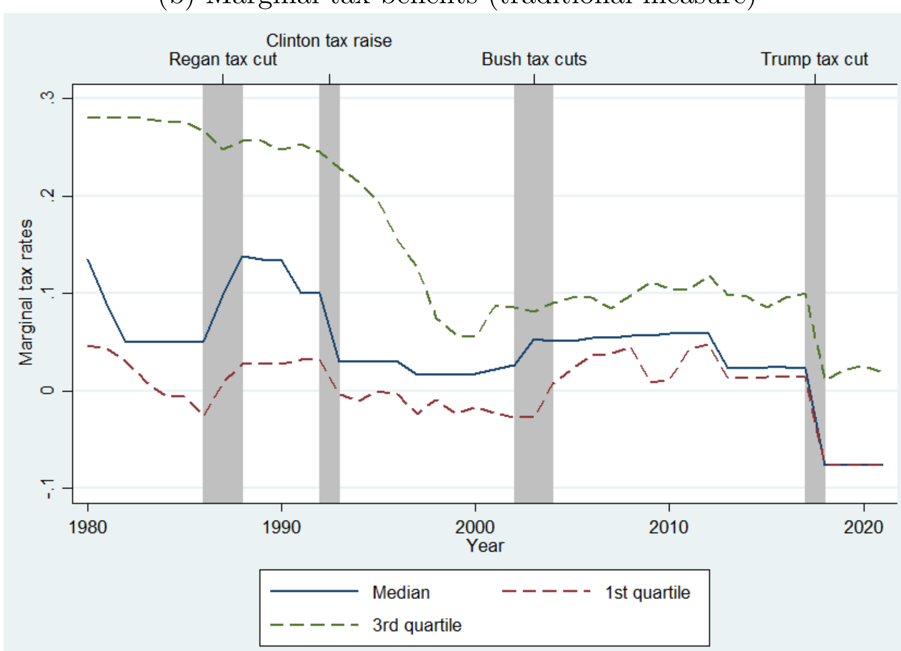
Figure 2.1 plots the time series of marginal tax benefits (new measure) and marginal tax rates (traditional measure) from 1980 to 2021 in the first, second, and third quartiles for observations with positive taxable income. In most years, the median marginal tax benefits are slightly below zero (between 0 and -10%), and the marginal tax rates are slightly above zero (between 0 and 10%). A typical profit-making firm faces no positive marginal tax benefits to issuing additional debt after adjusting for the personal tax consequences of leveraged recapitalization. Marginal tax benefits are left-skewed, while marginal tax rates are right-skewed. Under the traditional measure, firms with high earnings and relatively low interest expenses face high marginal tax rates even if adjusted with personal tax penalties by $(1-\tau_b)-(1-\tau_c)(1-\tau_e)$. However, since these firms usually have positive net payouts, issuing debt to raise future interest expenses leads to additional payouts with personal tax costs. Few firms face high marginal tax benefits after adjusting for such costs. In contrast, under the new measure, firms with low marginal corporate tax rates and positive net payouts face substantial personal tax costs for issuing debt through a leveraged recapitalization but little corporate tax savings. As a result, these firms face considerable negative marginal tax benefits for additional debt issuance.

Figure 2.1. Marginal tax benefits of debt 1980-2021

(a) Marginal tax benefits (new measure)



(b) Marginal tax benefits (traditional measure)



Notes. This figure shows the evolution of marginal tax benefits of debt from 1980 to 2021 in the first, second, and third quartiles for observations with positive taxable income. Panel (a) plots estimations of the new measure, that is, the marginal tax benefits for firms to raise a dollar of new perpetual debt after observing the current year's earnings. Panel (b) plots estimations of the traditional measure, that is, the marginal tax rates adjusted for personal taxes estimated, $(1-\tau_b)-(1-\tau_c)(1-\tau_e)$. Periods of tax reforms that substantially impact tax benefits of debt are marked in grey.

Several tax reforms have substantially affected firms' tax incentives for taking leverage. Ronald Reagan's tax reform in 1986 lowered the top federal personal income tax rates from 50% to 28% and top corporate income tax rates from 46% to 34%, effective in 1988. The larger cut in personal rates leads to a substantial increase in firms' marginal tax benefits from debt issuance. Bill Clinton's Omnibus Budget Reconciliation Act of 1993 raised the top personal rate to 39.6% and the top corporate rate to 35%, decreasing firms' marginal tax benefits from debt issuance. George W. Bush lowered the personal tax rates in his 2001 and 2003 tax reforms, with the top personal income rate dropping to 35%. After the reform, dividends are taxed at the rate of 15% instead of investors' personal income rates, which can be as high as 39.6% before the reform. The cut in personal income tax rates raised the marginal tax benefits, and the drop in dividend tax rates essentially shrank the difference between the first quartile and median marginal tax benefits in the new measure. In 2017, Donald Trump's tax cut substantially lowered the corporate tax rates from 35% to 21%. Marginal tax benefits dropped a lot. Since 2018 when the tax reform became effective, most firms are facing negative marginal tax benefits for debt issuance, even in the traditional measure. Given the low corporate rate relative to the personal rates, firms no longer have tax incentives to lever up, regardless of the purpose of debt.

2.3.3. The "kink" analysis

Here, I study the "kink" analysis, following Graham (2000) as a way to examine firms success in utilizing the tax shield of debt. Kinks are defined by the first level of interest expense, presented as multiples of actual interest expenses, where the marginal tax benefits drop by more than 50 basis points compared to the pre-financing level. Table 2.2 shows the summary statistics of estimated kinks. Estimates of the new measure show that over 90% of observations are already on the downward-sloping part of marginal tax benefits. The new measure implies a much smaller mass of underleveraged firms than the traditional measure. Only 4.67% of firms can double debt before tax benefits decline based on the new measure, compared to 20.36% based on the conventional measure. Most of the difference comes from correctly accounting for tax consequences of debt issuance on shareholders' personal taxes on equity income. Assessing the tax benefits from interest expenses in the future instead of current earnings further lowers estimations of unexploited tax benefits.

Table 2.2. Descriptive statistics for the kink analysis

	Ex-ante (1980-2021)				Ex-post (1980-2021)			
	new		traditional		new		traditional	
	obs	fraction	obs	fraction	obs	fraction	obs	fraction
kink <1	184,091	92.31%	149,046	74.73%	178,697	89.6%	139,997	70.2%
1 <= kink <2	4,451	2.23%	12,328	6.18%	5,631	2.82%	13,528	6.78%
kink = 2	1,571	0.79%	4,617	2.32%	2,018	1.01%	5,317	2.67%
kink = 3	2,573	1.29%	8,206	4.11%	3,507	1.76%	9,447	4.74%
kink >= 4	6,750	3.38%	25,239	12.66%	9,583	4.81%	31,147	15.62%
total	199,436	100%	199,436	100%	199,436	100%	199,436	100%

Notes. This table reports the number and fractions of observations with different levels of kinks, defined by the first increment of interest expense such that marginal tax benefits drop by at least 50 basis points. Ex-ante measures estimate the tax benefits for firms to raise a dollar of new perpetual debt after observing the current year earnings. Ex-post measures estimate the tax benefits for firms to add a dollar of interest expense directly to the observed current year earnings. New definitions are those taking the personal tax consequence of leveraged recapitalization into account, as discussed above. Traditional definitions are given by $(1-\tau_b)-(1-\tau_c)(1-\tau_e)$.

2.3.4. The zero-leverage puzzle

Strebulaev and Yang (2013) document a zero-leverage puzzle that about 1/5 firms have below 5% leverage, potentially leaving a substantial amount of money on the table by not taking advantage of the tax shield of debt. They did not adjust tax benefits of debt for personal tax penalty in their analysis due to the difficulty of observing firm-specific marginal personal rates. However, that effectively assumes the personal tax rates to be zero, as if tax-exempt institutions fully hold firms. However, a majority of shares are directly or indirectly owned by wealthy households who pay high tax rates on their income.[5] Here, I analyze whether firms with zero leverage or almost zero leverage can gain tax benefits from levering up after adjusting for personal taxes as described above.

Table 2.3 reports the summary statistics of marginal tax benefits for firms with below 5% book leverage to issue additional debt. Statistics for the subsample of observations with positive taxable income and the full sample are separately reported. The new definition considers all personal tax consequences of marginal debt issuance. The traditional definition adjusts tax benefits by Miller's formula. 46% of the observations have no taxable income to shield against. Firms with positive taxable income face a median marginal tax benefit of 3.7% by the traditional definition. Adjusting marginal tax benefits in the traditional way rationalizes the low leverage of over 25% firms with positive taxable income. Firms with positive taxable income face a median marginal tax benefit of -6.04% by the new definition. Additional tax cost from recapitalization rationalizes the low leverage of most firms that still seem underleveraged after the traditional adjustment. Only about

[5]Households may defer the tax payments if they hold securities, for example, through retirement plans.

10% of firms with positive taxable income and 5% of all (almost) zero-leverage firms face positive tax benefits from additional debt issuance.

Table 2.3. Marginal tax benefits for firms close to zero leverage

	positive taxable income		full sample	
	new	traditional	new	traditional
obs	33,733	33,733	62,552	62,552
mean	-9.7%	2.75%	-13.21%	-4.97%
std	12.5%	14.15%	12.16%	15.22%
min	-47.5%	-40%	-47.5%	-40%
1st quartile	-13.81%	-0.47%	-19.31%	-15.61%
median	-6.04%	3.7%	-11.24%	-3.94%
3rd quartile	-2.18%	9.68%	-4.54%	4.98%
max	24.97%	35.92%	24.97%	35.92%

Notes. This table shows the summary statistics of marginal tax benefits for observations with below 5% book leverage in the 1980-2021 sample period. Columns 2-3 show the summary statistics for observations with positive taxable income, and columns 4-5 show the summary statistics for the full sample. New definitions are those taking the personal tax consequence of leveraged recapitalization into account, as discussed above. Tradition definitions are given by $(1 - \tau_b) - (1 - \tau_c)(1 - \tau_e)$.

2.4. Leverage changes and marginal tax benefits

This section analyzes the relation between the measures of marginal tax benefits and firms' leverage changes in the data. If tax benefits and costs are among the key determinants of firms' leverage policies, this should be reflected in their decisions. Firms have a stronger incentive to raise leverage in the following year when they face higher marginal tax benefits of debt issuance, and vice-versa. The measure that is closer to the effective rates for firms' decision-making would explain the observed leverage changes better.

I use regressions of the following form to study how measures of marginal tax benefits explain firms' leverage changes in the following year.

$$\Delta MktLev_{i,t} = \beta_0 \tau^*_{i,t-1} + \beta_1 \Delta X_{i,t-1} + \alpha_{ind,t} + \epsilon_{it} \tag{2.3}$$

The dependent variable is change in firm's market leverage, defined by: Book value of debt (DLTT + DLC) / [Book value of debt (DLTT + DLC) + Market value of equity (CSHO * PRCC_F)]. The independent variables $\tau^*_{i,t-1}$ are lagged measures of marginal tax benefits. The control variables are standard ones commonly used in corporate debt research including profitablity (ROA), asset tangibility, firm size and market-to-book values, all in lagged changes form. $\alpha_{ind,t}$ represents the firm-year fixed effects.

Table 2.4 reports the regression results. When regressing changes in market leverage on the new and traditional measures of marginal tax benefits seperately, both measures have significant positive relations with changes in firms' market leverage in the following year. 1 percentage higher marginal tax benefits in the new (traditional) measure predicts about 2 (0.5) basis points higher increase in market leverage in the following year. That supports the view that firms take tax considerations as one of the key determinants for leverage decisions. When both measures are included together in the regression, the regression parameter for the new measure is still significantly positive but the sign of parameter for the traditional measure flips. That shows the new measure can better explain leverage changes than the traditional measure, and are better estimations of effective rates of tax consequences affecting firms' decisions.

Table 2.4. Leverage changes and marginal tax benefits

	Market leverage		
Marginal tax benefits	0.0229***		0.0418***
(new measure)	(0.0028)		(0.0044)
Marginal tax benefits		0.0057***	-0.0197***
(traditional measure)		(0.0022)	(0.0034)
ROA	-0.0102***	-0.0100***	-0.0102***
	(0.0025)	(0.0024)	(0.0025)
Tangibility	0.0369***	0.0370***	0.0366***
	(0.0107)	(0.0107)	(0.0107)
Size	0.0341***	0.0350***	0.0341***
	(0.0016)	(0.0016)	(0.0016)
Market-to-book	0.0000***	0.0000***	0.0000***
	(0.0000)	(0.0000)	(0.0000)
R^2	5.93%	5.89%	5.96%
Number of firms	12,950	12,950	12,950
Number of observations	127,564	127,564	127,564

Notes. This table reports the regression results for the relations between measures of marginal tax benefits and changes in firms' market leverage. Dependent variable is changes in market leverage, defined by $\frac{\text{Book value of debt } (DLTT+DLC)}{\text{Book value of debt } (DLTT+DLC)+\text{Market value of equity } (CSHO\times PRCC\_F)}$. Independent variables are lagged marginal tax benefits and lagged changes in ROA, asset tangibility, firm size and market-to-book values. The new measure of marginal tax benefits takes the personal tax consequence of leveraged recapitalization into account, as discussed above. The tradition measure is given by $(1-\tau_b)-(1-\tau_c)(1-\tau_e)$. All regressions are controlled for industry-year fixed effects. Standard errors are clustered at firm level.*** represents signigicance at the 1% level.

Such results are not surprising since the traditional measure is misspecified when the marginal debt issuance cannot be used to reduce equity issuance. If a firm is generating high profits that exceed its financing needs, additional debt issuance does not reduce equity issuance and thus shareholders' taxable equity income. Debt issuance only saves the firm corporate taxes at the cost of personal taxes priced in debt. In this case, the traditional measure, which always assumes savings of personal taxes on equity income, suggests this firm face a higher marginal benefit than another firm with lower profits and corporate tax rates. However, when the other firm is issuing debt as an alternative to equity, it can actually face higher marginal tax benefits than the high profit firm due to additional savings of personal on equity income. Such considerations are captured by the new measure of marginal tax benefits, explaining its stronger relation with the observed leverage changes.

2.5. Conclusion

This paper estimates a new empirical measure of marginal tax benefits for firms to issue new debt at the firm-year level. While many firms seem underleveraged according to the traditional measure, the new measure indicates that most firms see no benefits to levering up. The empirical evidence suggests that many firms having close to zero leverage is no longer a puzzle if we fully account for the personal tax costs for leveraged recapitalization. The new measure can better explain firms' observed leverage changes in the data.

CHAPTER 3

Undercutting the exchanges: private trading, fee competition, and price discovery at the market close (joint with Jiaheng Yu)

3.1. Introduction

Over recent years, the trading volume of US equities has been shifting to the market close, partly driven by passive investment strategies benchmarked against indices, which seek to trade at the market close price to minimize tracking errors (Bogousslavsky and Muravyev, 2021). The market close price of a stock is determined in a special call auction – the close auction, held by its listing exchange. From 2012 to 2018, the total trading volume executed in the close auctions has increased by 120% and started to account for more than 8% of the total trading volume. Meanwhile, the lack of competition drove up the close auction fees – NYSE's base rate has gone up by 16% and Nasdaq's by 60%, adding to the cost of benchmarking strategies. [1]

"Guaranteed close", offered by investment banks like Goldman Sachs, executes clients' orders at the market close price set by the close auctions yet charges a lower transaction fee. Meanwhile, it allows the banks to trade and profit on the order flow information – after pairing buyers with sellers, the banks hedge the imbalance and trade on their principal

[1] See Tier F under Execution Fees for the Nasdaq Closing Cross in the Nasdaq fee schedule available at [link]; and Liquidity Indicator 7 in the NYSE fee schedule available at [link].

accounts. Trading volume through "guaranteed close" has reached almost 30% of that through close auctions in 2018. [2] Albeit this large trading volume, it is understudied how "guaranteed close" affects the formation of close prices, which are arguably the most important prices, widely used in net asset values, margin accounts, numerous financial contracts and risk metrics. Importantly, if "guaranteed close" makes close prices less informative, that is, farther from the fundamental values, it imposes higher costs for passive investment strategies that benchmark and trade at close prices, and undermines the role of stock market price discovery in allocating resources (Bakke and Whited, 2010; Bond, Edmans, and Goldstein, 2012). Hence, we undertake the task of formally studying the impact of "guaranteed close" on price discovery, both empirically and theoretically.

It is ex-ante unclear how "guaranteed close" might affect price discovery. In fact, close auctions are considered robust mechanisms in generating the close prices, and off-exchange venues like "guaranteed close" that siphon trading activities have raised public concerns. Executives at NYSE and Nasdaq suggested that "... if more trading moves to banks, it will make close prices less trustworthy...". Besides these incumbents, Credit Suisse Trading Strategy suggests in a market commentary that "...changes to end-of-day trading dynamics could indicate that there has been some impairment to the price discovery process". In comments submitted to SEC [3], 41 listing companies (85% of the respondents), including PayPal and FedEx, voiced against fragmentation and disruptions

[2]Based on our calculation. Also documented by Credit Suisse Trading Strategy, WSJ (2018) and SEC report: Securities Exchange Act Release No. 34-80683 (May 6, 2017).

[3]To be precise, these are comments to the proposal of Bats Market Close, a closing match process that supplements NYSE and Nasdaq close auctions. The proposed process is similar to guaranteed close, charging lower fees and using prices set by close auctions. Notably, the public has general opposition to fragmentation and disruptions at market close.

to the close auctions, suggesting that "... [fragmentation] will increase volatility and decrease precision in closing prices."

Our investigation is also well suited to speak to a more general question: how does trading on (retail) order flow affect price discovery? The theoretical literature on dual trading (Fishman and Longstaff, 1992; Röell, 1990; Sarkar, 1995) and back-running (Huddart et al., 2001; Yang and Zhu, 2020) shed lights on this question. It is, however, empirically challenging due to the lack of data on exogenous shocks to trading on order flow activities. This question becomes eminent as e-brokers like Robinhood flock to charge zero commission fees yet sell order flow data to sophisticated institutions for them to exploit and trade on. As will be discussed soon, our empirical setting provides a nice opportunity to study this question in the context of today's US brokerage industry.

We provide quasi-experimental evidence that "guaranteed close" improves price discovery. To start our analysis, we measure each stock's trading volume via "guaranteed close" by the volume of off-exchange trades between 4:00 p.m.–4:10 p.m. EST executed at the official close price using the NYSE Trade-and-Quote (TAQ) millisecond-level data.[4] We show that ETF and index fund ownership strongly correlate with the "guaranteed close" volume, indicating that "guaranteed close" is mainly used by passive investment strategies to save expenses on transaction costs.[5] We measure the informativeness of the close price by the "closeness" between the close price and the next day's open price. Specifically, we compute the mean of squared overnight return for individual stocks at

[4]SEC DERA (2017) confirmed that these trades are almost surely from off-exchange market-on-close orders. "Guaranteed close" is the major venue that executes off-exchange market-on-close orders.

[5]This echoes anecdotal evidence that "guaranteed close" is used by index-fund managers including Vanguard Group and BlackRock Inc. (WSJ, 2018)

a monthly frequency as a measure of close price's informativeness. This measure corresponds to the mean squared error measure of price informativeness used in the theoretical literature.

Next, we use the NYSE close auction fee cut in January 2018 as a quasi-experiment and use a standard difference-in-differences strategy to estimate the effects of banks' trading on order flow activities. In the fee cut, NYSE reduced the Tier 1 and Tier 2 rate charged to broker-dealers executing high trading volumes in the close auction significantly by around 40%, but kept rates for the remaining broker-dealers almost unchanged, with the non-tier rate reduced by only 9%. This fee cut is part of NYSE's effort to draw trading activities back to the close auctions, and this incurred differential impact on different stocks. We find that stocks ex-ante more heavily traded at "guaranteed close", and those with higher passive ownership are more exposed to the fee cut – the trading volume of these stocks at "guaranteed close" significantly dropped relative to the remaining stocks.

There are at least two reasons why the fee cut for broker-dealers passes through to individual stocks in such a differential manner. First, ETFs and index funds may actively choose the venue at which they trade and are sensitive to transaction fees. The fee cut induced them to actively choose to trade with broker-dealers that operate in the close auction. Their large trading volume easily helps broker-dealers reach the Tier 1 and Tier 2 thresholds, allowing them to take the lower transaction fees effectively. Second, institutional investors may delegate the venue choice entirely to their broker-dealers. They rely on broker-dealers to trade (Di Maggio, Egan, and Franzoni, 2021), build long-term relationships with their brokers, and concentrate their order flow with a relatively small set of broker-dealers (Goldstein, Irvine, Kandel, and Wiener, 2009). In that case, broker-dealers would have a different yet relatively stable client base. Some provide services mainly to passive investors like ETFs and index funds, while others provide services mainly to other types of investors. Broker-dealers' trading on behalf of passive investors is more exposed to the fee cut since they are more likely to reach the Tier 1 and Tier 2 thresholds. Once they go back to the close auctions, the stocks they usually facilitate to trade, that is, those with high passive ownership, will experience a larger drop in trading volume at "guaranteed close". These stocks are also ex-ante heavily traded at the "guaranteed close". Under both cases, and to the extent that banks trade on the order flow data they receive in the "guaranteed close", the fee cut is a plausibly exogenous shock to trading on order flow activities.

For each stock, we calculate the trading volume executed at "guaranteed close" as a fraction of the trading volume executed at the close price and term it as the "guaranteed close fraction". Trading volume executed at the close price includes those executed in the close auction and those in the "guaranteed close". We sort all NYSE stocks by their average "guaranteed close fraction" before the NYSE fee cut. The treated group consists of stocks that rank in the top 50%. The control group consists of the remaining stocks. Our difference-in-differences estimation finds that the NYSE fee cut significantly decreased the informativeness of the close prices for the treated stocks, relative to the control stocks, by 15.7% compared to the sample mean. Indeed, treated stocks may have vastly different characteristics than the control stocks – for example, they have larger market capitalization and higher passive ownership. To alleviate this concern, we control for a wide range of variables: stock fixed effects, time fixed effects, market cap, total trade volume, intra-day volatility, measures of total retail volume and institutional volume, after-market-close volume/total volume, close auction volume/total volume, and overnight betas of individual stocks. We also verified the parallel trend assumption: treated stocks are no different in price informativeness from control stocks before the NYSE fee cut. In addition, we show that intra-day volatility and quoted bid-ask spread during market hours did not respond to the NYSE fee cut shock. That means our result does not merely reflect structural changes in the general trading activities of treated stocks. Our result is also robust to different measures of price informativeness, such as the median of squared overnight returns and the mean absolute value of overnight returns.

We conduct additional exercises to further the robustness of our difference-in-differences estimation results. First, we adopt a different rule to designate the treatment group. Specifically, we rank stocks by their ETF and index fund ownership before the NYSE fee cut. The treated group consists of stocks that rank in the top 50%. The control group consists of the remaining stocks. A similar difference-in-differences estimation suggests that NYSE fee cut decreased the informativeness of treated stocks by 14.3%. Second, our price informativeness measure takes the next day's market open price as the fundamental value to which we compare the market close price with. Although a natural measure, the next day's market open price also incorporates overnight information, and it is possible that the strength of overnight information of treated and control stocks coincidentally changed after the NYSE fee cut. To alleviate this concern, we exclude data in the 3-day window around earnings announcement days and conduct the same difference-in-differences estimation, and arrive at similar results. Third, to further address the concern that treated and control stocks have different characteristics, we conduct a matched sample difference-in-differences estimation, using control group matched with treated group on pre-treatment values of market cap, trading volume, intra-day volatility, and overnight beta. The results are similar.

Our estimated effect of NYSE fee cut on mean absolute value of overnight returns shed light on how the fee cut affects the profit to investors of index funds and ETFs. The difference-in-differences estimated increase in mean absolute value of overnight returns is 5 bps. The relative fee cut between treated and control group is 2 bps. The impact on investor profits depends on the probability that a stocks overnight movement is against the index funds trade, that is, the stock price goes up yet the index fund sold it at market close, and that the stock price goes down yet the index fund bought it at market close. For reasonable estimate of this probability, back-of-envelop calculation suggests that the benefit of the decline in fee is outweighed by the cost of increasing liquidity pressure in the close price for trades that are executed in the close auction. Each passive investor that chooses to go to the close auction for a lower fee ignores her impact on the price, and such an externality in the aggregate makes passive investors worse off.

How could the "guaranteed close" improve price discovery? Insights from existing studies that suggest dark pools[6] can improve price discovery (Zhu, 2014) cannot be directly applied to "guaranteed close". In Zhu (2014)'s model, since dark pools do not absorb excess order flows, informed traders, who are more likely to trade in the same direction as each other, face a higher execution risk in dark pools relative to uninformed traders. Hence dark pools would concentrate the informed traders on the exchange and improve price

[6]Dark pools are equity trading systems that do not publicly display orders. They typically passively match buyers and sellers at exchange prices, such as the midpoint of the exchange bid and offer.

discovery. However, there is no execution risk in "guaranteed close". Also importantly, while dark pools typically passively match buyers and sellers and do not take their own positions, banks actively trade on the order flow information they received in "guaranteed close", which might also affect price discovery.

To explain our finding that "guaranteed close" improves price discovery, we build a model based on the single-period model in Kyle (1985). In our model, traders explicitly choose between two trading venues – that is, a bank and a market maker, based on transaction costs. The bank infers information from the net total orders it receives and trade. It trades in the same direction as the net total orders, when the proportion of informed orders in the orders sent to the bank is higher than those sent directly to the market maker. And it trades in the opposite direction otherwise. The bank's trading activity amplifies the informed orders eventually received by the market maker and hence improves price informativeness.

Uninformed traders and informed traders all have unit demand for a single asset with an uncertain liquidation value v. Each informed trader receives a signal about v and trades according to the signal. Each uninformed trader has to trade after receiving a liquidity shock. Both the market maker and the bank accept only market orders. Trading with the bank incurs a convenience cost, which is heterogeneous among traders. The market maker, being competitive, sets the price p. A trader can submit his order either to the market maker or the bank. While both venues execute his order at p, the transaction fees are different – the market maker charges φ_m per unit of asset traded and the bank charges φ_b. φ_m is exogenous and φ_b is chosen by the bank to maximize profit. While the bank executes orders for the traders, it also submits orders on its own account to the market maker for profit. The transaction fees and the distributions of traders' features are publicly known.

We solve the model in closed form, and find a linear Nash equilibrium where the net total orders of informed traders is linear in the asset's value v, the bank's trading strategy is linear in the net total orders it receives, and the market maker sets p as a linear function of the net total orders he receives. Under mild conditions, the bank participates with a fee $\varphi_b < \varphi_m$. The informativeness of p is higher than in an equilibrium without the bank. This is because the bank can trade profitably based on the net total orders received so long as the proportions of informed orders relative to uninformed orders are different between the two venues. The bank's trading activity amplifies the informed orders eventually received by the market maker, making the close price more informative. We decompose the effect of φ_m on the informativeness of p into two components – a volume effect and a ratio effect. The volume effect operates through the trading volume executed by the bank, and find that it is positive, meaning that higher trading volume at the bank leads to higher price informativeness. This is consistent with our empirical finding. The ratio effect instead operates through the ratio of informed relative to uninformed orders received by the bank. Finally, we find that dual trading of the bank unambiguously decreases the expected profit (before fees) of informed traders, and has a mixed effect on the expected profit (before fees) of liquidity traders.

Our empirical results and model are well-intended to study the effect of trading on order flow in today's US brokerage industry. Today's brokers offer assorted venue choices with different transaction fees. Some brokers like Robinhood charge zero transaction fee yet sell order flow data to sophisticated institutions for them to trade on. The combination of Robinhood and its partners who trade on order flow data is comparable to the "guaranteed close". Meanwhile, other brokers charge a higher transaction fee yet do not actively trade on order flow for profit, which is comparable to the "close auction". The brokerage firm Interactive Brokers even simultaneously offers two options: IBKR Lite and IBKR Pro. "If it is IBKR Lite with zero commissions ,..., we send them off to a market maker,..., and there is payment for order flow that comes back and you may not get as good of an execution ... If it is IBKR Pro, you will get better execution."(Steve Sanders, executive vice president of Interactive Brokers, link). In our empirical setting, designated market makers clear the market in a single auction and set the price. This fits nicely with the Kyle (1985) framework that has been successfully applied to understand the formation of prices generally. In our model, "guaranteed close" improves price discovery as long as the fraction of informed orders relative to uninformed order sent to the bank is different than that to the market maker. Without stipulating the reasons why the fractions of informed orders might be different, our model offers a powerful partial equilibrium result.

Let us finally caution that our model is stylized, with exogenous noise trading, and no entry of banks, among other assumptions. A dynamic model with multiple exchanges and multiple broker-dealers would be more realistic if extrapolating our result to other trading scenarios. In addition, our model is only meant to capture one aspect of "guaranteed close", namely its effect on price discovery. The other side of the coin, namely, the liquidity provision feature, especially whether it still provides liquidity when market conditions are volatile, is not studied. Future research may shed more light on these issues.

Literature Review

Our paper is related to three strands of literature. First, we add to the recent discussion on how the growth of passive investment strategies affects the market. Although market closures by themselves can generate endogenous time-variation in trading activity and price movements (Hong and Wang, 2000), Bogousslavsky and Muravyev (2021) find that the influx of ETF and index fund trades are key determinants of the volume at market close in recent years, and they adversely affect the informativeness of the close prices. Ben-David, Franzoni, and Moussawi (2018) find that ETF ownership is associated with higher volatility and more reversals for the index constituents. Baltussen, van Bekkum, and Da (2019) find that the growth of passive investment is associated with a decline in index return autocorrelation. Lines (2022) finds that when market volatility rises, portfolio tracking error also rises, which leads portfolio managers to rebalance their portfolios towards benchmark stocks, and this generates price effects. Baldauf, Frei, and Mollner (2021) study the manipulation of prices at market close. They build a model of financial contracting between a client, who wishes to trade a large position, and her dealer. Because of agency problem, market-on-close order is not the optimal contract for trading. Our paper focuses on the costs from transaction fees and price pressures, that are adversarial to passive investment strategies. We hope to inform policy attempts that design market structure to accommodate the growth of passive investment strategies.

Second, our paper contributes to the literature on dual trading, that is, broker-dealers strategically using customers' order flow information to trade on their own accounts. On the empirical side, Chakravarty and Li (2003) use proprietary audit trail transaction data to study dual trading in futures markets, and find that dual traders trade merely to supply liquidity and manage inventory, rather than trading against the customers.

Barbon, Di Maggio, Franzoni, and Landier (2019) find that broker-dealers intermediating large stock portfolio liquidations spread order flow information to their clients. Using data provided by Robinhood, Kothari, Johnson, and So (2021) find that payment for order flow has saved retail investors unnecessary trading commissions, and improved the execution quality. We provide quasi-experimental evidence that trading on order flow can improve price discovery.

On the theory side, Röell (1990) builds a model where the broker observes only the trades of uninformed trader and trade on them. In Fishman and Longstaff (1992), the broker has private information about whether his customer is informed or not, and allow the customer and the dual-trading broker to trade at different prices. Sarkar (1995) finds that dual trading has no impact on discovery in a fully-revealing equilibrium, although it decreases net profits of informed traders and increases the utility of uninformed traders. Yang and Zhu (2020) and Huddart et al. (2001) study the strategies of informed traders when there are back-runners, who partly infer informed traders' information from their order flow and exploit it in subsequent trading. The informed traders counteract back-runners by randomizing their orders (unless back-runners' signals are too imprecise), but back-runners unambiguously improve price discovery. These analyses are different from ours in terms of the economic questions and modeling approaches. In our model, the bank observes only the net orders and partially infer information from order flows. Also, we explicitly model the venue choices, and all the trading happens in one period and the price is set only once.

Finally, our paper is related to the literature on dark pools and alternative trading venues. Zhu (2014) shows that adding a dark pool alongside the exchange can improve price discovery due to the dark pool's execution risk. Buti, Rindi, and Werner (2017) find that dark pools may have adverse effects on market quality, since dark pools reduce the number of limit orders that provide liquidity on the exchange. Ernst, Sokobin, and Spatt (2021) find that market participants learn from the publication of off-exchange transactions, and the off-exchange orders are informationally-motivated and contribute to price discovery. Other models on trading venue choice include Hendershott and Mendelson (2000) and Ye (2010). Their models either do not model asymmetric information, or do not allow all the agents to freely select venues, and do not consider the commission fee difference. Chen and Duffie (2021) find that market fragmentation and exchange competition could lead to improvement in price discovery when all exchange prices are taken together. Brogaard and Pan (2021) provides evidence that that dark pool trading leads to greater information acquisition.

The remainder of this paper is organized as follows. In Section 2, we present the institutional background of close auctions on the primary exchanges and "guaranteed close" of the banks. Section 3 exhibits quasi-experimental evidence that "guaranteed close" improves price discovery. Section 4 introduces our model of dual trading and price discovery at market close. Section 5 concludes.

3.2. Institutional Background

3.2.1. Close auction

In this section, we introduce the mechanism of NYSE's close auction, whose characteristics are to a large extent shared by other exchanges like NYSE Arca and Nasdaq.[7]

Several types of orders can be used in NYSE's close auction, with the most common being market-on-close (MOC) and limit-on-close (LOC) orders. An MOC order is an unpriced order to buy or sell a security at the close price and is guaranteed to receive an execution. An LOC order sets the maximum price an investor is willing to pay, or the minimum price for which an investor is willing to sell. An LOC order priced better than the final close auction price is guaranteed to receive an execution. As shown in Table C.1 in the appendix, 65% of the orders in NYSE close auction are MOC orders and 14% are LOC orders.[8] The predominant use of MOC orders appears to be the consequence of benchmarking strategies conducting trades at market close price, regardless of what the price will be.

From 6:30 am in the morning, MOC and LOC orders can be entered. Existing MOC and LOC orders can be canceled until 3:50 pm. At 4:00 pm, the regular session trading ends and the close auction commences. The method for determining the close prices follows two principles: (1) maximize the number of shares that can be executed in the close auction; (2) minimize the difference between the close price and a reference price if multiple close prices satisfy principle (1). The auction effectively aggregates the supply

[7]See Appendix A of Bogousslavsky and Muravyev (2021) for a detailed summary of Nasdaq's close auction mechanism, as well as NYSE's. Also, see "NYSE Open and Closing Auctions Fact Sheet", 2019 [link], "NYSE Arca Auctions Brochure", 2019 [link], and "Nasdaq Open Close Quick Guide", 2019 [link]

[8]Another 18% are Closing D Orders, a special order accessible only to NYSE floor brokers.

and demand curve constituted by the MOC and LOC orders, and the transaction price and trade volume is determined by the intersection of the two curves.

The Designated Market Makers (DMMs) play an important role in the close auction. They set the closing price at a level that satisfies all interest that is willing to participate at a price better than the close auction price, and supply liquidity as needed to offset any remaining auction imbalances that exist at the closing bell. That means market-on-close orders are guaranteed to be executed.

3.2.2. "Guaranteed close" service

Investment banks such as Goldman Sachs, Morgan Stanley, Credit Suisse Group AG, and UBS Group AG, started a "guaranteed close" service around 2016.[9] Investors looking to buy or sell shares of a stock can get a guarantee from the bank to execute their orders at the close price set on the corresponding primary exchange, where the stock is listed. As an investor, using "guaranteed close" is equivalent to sending a MOC order to the close auction in execution outcomes, but paying a lower fee. People familiar with the matter told us that Goldman Sachs recently cut the fee charged to broker-dealers to zero (although buy-side clients still pay a fee).

At 4:00 p.m., the bank pairs the buyers with the sellers of the stock. For the unmatched orders, it can either send them to the exchange or take the other side of the trade itself, storing the extra shares or short interest on its books overnight. The banks trade alongside the client orders for profit. This is one way to cover the bank's cost of providing liquidity,

[9] This is basd on multiple sources. We do not, however, have a comprehensive list of the venues that conduct "guaranteed close". For the top ten alternative trading systems by total volume (during both regular hours and market close), see Table 4 in "Staff Report on Algorithmic Trading in U.S. Capital Markets", 2020, SEC.[link]

and is documented by the following excerpts from the documentation of "guaranteed close" Morgan Stanley and Goldman Sachs sent to their clients.

Morgan Stanley:[10]

> "When we accept an order for execution on a guaranteed benchmark basis (for example, a guaranteed opening, closing, volume weighted average price or other guaranteed transaction), we will typically attempt to offset the risk incurred as a result of such guarantee by transacting in the market on a principal basis, or accessing internal liquidity sources, in the benchmark security or a related instrument, although we may choose not to perfectly hedge our exposure."

Goldman Sachs:[11]

> "We offer client facilitation services, which are typically used by clients to obtain liquidity or a guaranteed execution price. When you use our client facilitation services, we may also effect transactions as agent, as principal (including trading as a market maker or liquidity provider to other clients and trading to manage risks resulting from client facilitation activities), or in a mixed capacity."

Also, Morgan Stanley claims to split the profit from dual trading with the client:

[10] See "A Message to Morgan Stanley's U.S. Institutional Equity Division Sales & Trading Clients regarding U.S. Equity Order Handling Practices", 2019, [link]

[11] See "Cash Equities Order Handling Procedures of Goldman Sachs (Asia) L.L.C.", 2019,[link]

> "In accordance with market standards and best practices, we strive for allocations between client and principal orders that are fair and equitable.[...]Challenges presented by the current market structure and limitations of certain market centers and trading system may in some cases render a precisely even split impracticable."

Who are the users of the "guaranteed close"? Anecdotal evidence suggests the "guaranteed close" is used by index-fund managers including Vanguard Group and BlackRock Inc., as well as some smaller broker-dealers (WSJ (2018)). In general, institutional investors and broker-dealers are primary users of alternative trading systems.[12] In the next section, we present evidence that stocks with higher ETF/index fund ownership are more heavily traded at the "guaranteed close".

"Guaranteed close" is different from dark pools in that it has no execution risk, while in the latter matching depends on the availability of counterparties and some orders from the "heavier" side of the market will fail to be executed (Zhu, 2014). Also, it is also in nature different from the recently approved Cboe Market Close program providing a lower-fee venue to trade at the market close price set by NYSE and Nasdaq. In Cboe Market Close, traders can enter, cancel or replace MOC orders only before 3:35pm. After that, the system would match for execution all buy and sell MOC orders entered with execution priority given based on time-received. But any remaining balance of unmatched shares would be cancelled and returned to the traders. That is, Cboe Market Close only pre-matches some non-price-forming orders.[13]

[12] See Section IV in "Staff Report on Algorithmic Trading in U.S. Capital Markets", 2020, SEC.[link]

[13] See WSJ article "SEC Decision on 4 p.m. Closing Trades Deals Blow to NYSE, Nasdaq " at [link], and SEC Release No. 34-88008 at [link].

3.3. "Guaranteed Close" and Price Discovery: Evidence from NYSE Fee Cut

In this section, we empirically study the relationship between "guaranteed close" and the informativeness of market close price. We exploit the quasi-experimental setting of the NYSE close auction fee cut in January 2018 and use a difference-in-differences approach to establish causal evidence.

3.3.1. NYSE close auction fee cut

In January 2018, the NYSE reduced the close auction fee for market-on-close (MOC) orders, which was seen as an attempt to keep clients from choosing the banks' low-cost "guaranteed close" service.[14] Indeed, the NYSE only reduced the fee for MOC orders but kept the fee for limit-on-close (LOC) orders unchanged. The NYSE's claimed intention is to encourage higher volumes of MOC orders at the close. This fee cut was given with short notice. On December 21, 2017, the NYSE announced the plan of fee changes intended to be effective January 2, 2018. On January 8, 2018, NYSE filed with the SEC about the change in close auction fees.[15]

In NYSE's close auctions, the amount of fee broker-dealers need to pay for the MOC orders depends on the trading volume – the per-share fee is lower for a higher trading volume. Specifically, there are three tiers. Tier 1 rates would be available for a broker-dealer who in the prior three billing months executed (1) an ADV (average daily volume) of MOC activity on the NYSE of at least 0.45% of NYSE CADV (consolidated average daily volume), (2) an ADV of total close activity (MOC/LOC and executions at the close)

[14] Alternatively, the fee cut may be seen as a reaction to the threat of CBOE's entry into the close auction [link], which was awaiting SEC decision back in 2018.

[15] See SEC No. 34-82563 [link]. Also, see NYSE trader update [link].

on the NYSE of at least 0.7% of NYSE CADV , and (3) whose MOC activity comprised at least 35% of the its total close activity. Tier 2 rates would require a lower ADV of MOC activity and ADV of total close activity. Those who don't meet the requirements for Tier 1 and Tier 2 rates are subject to the Non-Tier rate.

The fee cut reduced the Tier 1 rate and Tier 2 rate significantly, but left the Non-Tier rates almost untouched. Specifically, Tier 1 rate is reduced from $0.0007 to $0.0004 (a 42.9% drop). Tier 2 rate is reduced from $0.0008 to $0.0005 (a 37.5% drop). But Non-Tier rate is reduced from $0.0011 only to $0.0010 (a 9% drop).

Given the differential changes for different tiers, we expect the fee cut to have differential impact on individual stocks. There are at least two reasons why the fee cut for broker-dealers pass through to individual stocks in a differential manner. First, ETFs and index funds may actively choose the venue at which they trade and are sensitive to transaction fees. The fee cut induced them to actively choose to trade with broker-dealers that operate in the close auctions, this is because their large trading volume easily help broker-dealers reach the Tier 1 and Tier 2 thresholds, allowing them to effectively take the lower transaction fees. Second, institutional investors may delegate the venue choice entirely to their broker-dealers – this is a reasonable assumption as institutions rely on broker-dealers to trade (Di Maggio, Egan, and Franzoni, 2021), form long-term relationships with their brokers, and concentrate their order flow with a relatively small set of broker-dealers (Goldstein, Irvine, Kandel, and Wiener, 2009). In that case, broker-dealers would have different yet relatively stable client base. Some provide services mainly to passive investors like ETFs and index funds, while others provide service mainly to other types of investors. Broker-dealers trading on behalf of passive investors are more exposed to the fee cut since they more likely reach the Tier 1 and Tier 2 thresholds. Once they went back to the close auctions, the stocks they usually facilitate to trade, that is, those with high passive ownership, will experience a larger drop of trading volume at "guaranteed close". These stocks are also ex-ante heavily traded at the "guaranteed close". Under both cases, and to the extent that banks trade on the order flows data they receive in the "guaranteed close", the fee cut is a plausibly exogenous shock to trading on order flow activities. Stocks that are ex-ante heavily traded at the "guaranteed close", and those with higher passive ownership, are more exposed to the fee cut.

3.3.2. Data and descriptive findings

Data construction. Our main data source is the NYSE millisecond-level trade and quote data (TAQ) spanning from 2012 to 2019. We also leverage the WRDS Intraday Indicator Dataset (WRDS IID) built from the TAQ data. Our sample contains common stocks listed on NYSE and Nasdaq, with a price greater than $5 and a market capitalization greater than $100 million at the end of a month. Our main results come from the difference-in-differences analysis, which uses the sample of 1,217 NYSE stocks, spanning from January 2015 to December 2019. Table 3.1 reports the summary statistics of all the variables we used in this sample.

Table 3.1. Summary statistics: NYSE sample

	N	Mean	SD	10th	50th	90th
MSE ($\times 10^4$)	69,273	0.56	0.54	0.09	0.39	1.30
MSE1 ($\times 10^4$)	69,186	0.76	0.73	0.12	0.52	1.74
MSE2 ($\times 10^4$)	69,100	0.54	0.50	0.09	0.38	1.22
MSE3 ($\times 10^4$)	69,159	0.54	0.51	0.09	0.38	1.24
Median SE ($\times 10^4$)	69,273	0.21	0.28	0.03	0.11	0.51
MAD (percent)	69,273	0.51	0.26	0.23	0.44	0.87
Index fund ownership	69,294	0.07	0.04	0.00	0.08	0.11
ETF ownership	69,294	0.06	0.05	0.00	0.07	0.13
Off-ex MOC volume/close price volume	69,291	0.11	0.09	0.00	0.09	0.25
Volatility ($\times 10^6$)	69,288	0.87	2.34	0.03	0.15	1.72
Market cap ($ bil.)	69,290	13.68	27.13	0.32	3.79	35.27
Total volume ($1,000$ shares)	69,293	1606.41	2447.13	56.05	690.57	4185.92
Total retail volume ($1,000$ shares)	69,258	92.52	170.06	4.78	29.89	238.30
Total volume of trades $\geq$ $20K in value ($1,000$ shares)	69,027	415.31	773.41	11.31	123.90	1069.77
Total volume of trades $\geq$ $50K in value ($1,000$ shares)	68,457	257.97	485.67	9.28	76.26	661.82
After-close volume/total volume	67,652	0.04	0.04	0.01	0.03	0.09
Close auction volume/total volume	69,291	0.08	0.05	0.01	0.08	0.15
Overnight beta	68,769	0.91	0.50	0.24	0.91	1.56
Spread (percent)	69,292	15.56	15.75	3.50	10.16	33.20

Notes. This table reports the summary statistics of all the variables used in this paper. The sample consists of 1,217 NYSE stocks, and spans from 2015m1 to 2019m12. Variables with daily observations are aggregated at a monthly frequency by calculating averages. MSE, MSE1, MSE2, MSE3 are mean squared error measures of price informativeness, calculated from daily data. Specifically, they are the monthly average of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open5m}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close5m}})^2$, and of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close15m}})^2$, respectively. Median SE is the monthly median of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$. MAD is the monthly average of $|\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}}|$. Index fund ownership and ETF ownership are the fraction of outstanding shares held on average in 2016 by index funds and ETFs, respectively. Close price volume is volume traded at close price (i.e., close auction volume + Off-ex MOC volume). Volatility is trade-based intrayday volatility during market hours. Total volume is total trade volume during market hours. Total retail volume is total volume of retail trades during market hours. After-close volume is trade volume after market close and before next day's market open. Overnight beta is individual stock's overnight return's loading on overnight market return, estimated quarterly using CAPM regression. Spread is time-weighted percent quoted spread during market hours.

The close price, open price, and close auction volume for each stock and each day are from WRDS IID. The close auction volume is measured by TAQ trades with the sale condition of 6 (Closing Print), that occurs on a stock's primary listing exchange (SEC DERA, 2017) – for example, an NYSE-listed stock's close volume executed in NYSE.[16] The close prices are the recorded transaction prices of these trades. The market open prices are measured by the recorded prices of the trades with the sale condition of O (Market Center Opening Trade). For each stock each day, we also calculate from TAQ the volume-weighted average price in the last 5 minutes, $p_t^{close5m}$, in the last 15 minutes, $p_t^{close15m}$, and in the first 5 minutes, p_t^{open5m}, of the regular session. For these calculations, we exclude invalid or erroneous trades that were later canceled or changed.

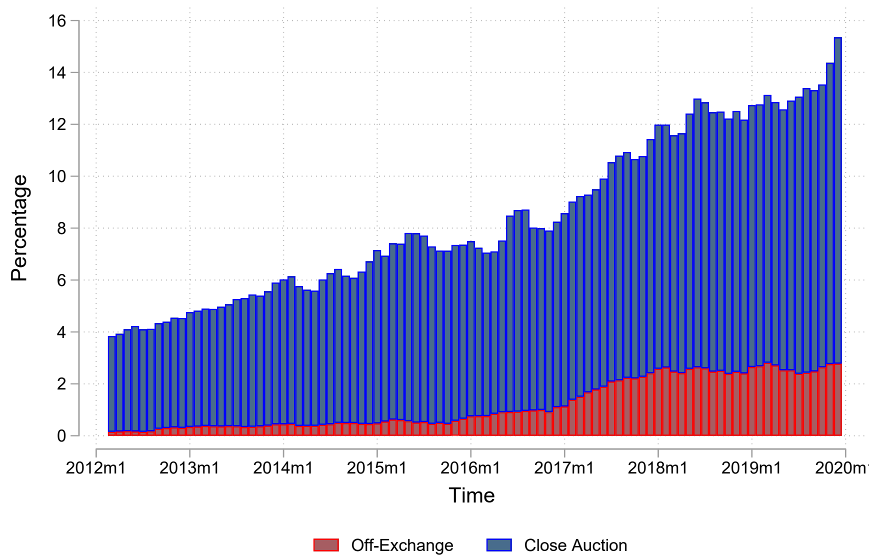
We proxy the trade volume via "guaranteed close" service by the trade volume using off-exchange market-on-close (MOC) orders, given that "guaranteed close" is anecdotally the major venue accepting off-exchange market-on-close orders. To be precise, we use the term "off-exchange MOC volume" in the formal analysis. To measure the off-exchange MOC volumes, we consider all the trades from TAQ that are not cancelled or corrected and occur between 4:00 p.m.–4:10 p.m. EST at the official market close price determined by the close auction. This off-exchange MOC volume has been validated against two regulatory datasets with more detailed information – the FINRA Trade Reporting Facility data, and the FINRA-provided Audit Trail data, to ensure that it is from MOC orders.[17] The two regulatory datasets identify off-exchange executions by venue and trace the executions back to the original orders. Our off-exchange MOC volume would be almost identical if measured by the regulatory datasets. As shown in Figure 3.1, the trading

[16]NYSE stocks technically can also be traded in Nasdaq closing crosses, which is yet empirically rare.
[17]See SEC report SEC DERA (2017).

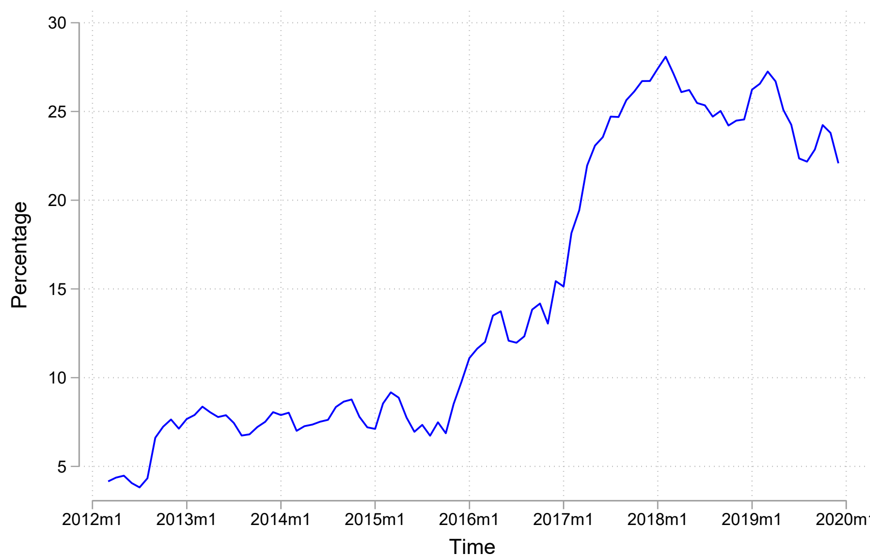
volume in “guaranteed close” has risen sharply from 2016, and reached almost 30% of the trading volume in close auctions.

Figure 3.1. Trade volume of S&P 500 stocks in "guaranteed close" and close auctions

(a) Volume in "guaranteed close" and close auctions



(b) Volume in "guaranteed close"/ volume in close auctions



Notes. Panel (a) shows the trade volume in "guaranteed close" and trade volume in close auctions as a percentage of total trade volume for S&P 500 stocks from 2012m1 to 2019m12. Panel (b) compares trade volume in "guaranteed close" with trade volume in close auction. In both panels, we first aggregate the volumes of each stock to monthly observations and calculate the percentages, then smooth the time series by taking three-month moving average.

ETF ownership data is obtained from ETF Global. Index fund ownership data is obtained from the CRSP Mutual Fund database. We closely follow Ben-David et al. (2021)[18] and Dannhauser and Pontiff (2019) to identify passive index funds in the CRSP Mutual Fund database. Our sample contains 918 index funds in 2016.

For other variables, we obtain market capitalization and earnings announcement days from CRSP. Measures of volatility, liquidity, and retail and institutional order flows are from WRDS IID. They include the trade-based intraday volatility during market hours, total trade volume during market hours, total retail trade volume following Boehmer et al. (2021), total volume of trades $\geq$ 20K in value, and total volume of trades $\geq$ 50K in value. The last two variables are proxies for institutional trades. Lee and Radhakrishna (2000) show 53% of institutional trades are above $20,000 in value; Bhattacharya et al. (2007) and Shanthikumar (2003) use $50,000 dollar value-based cutoff. When trades with value exceeding these cut-offs are interpreted as institutional trades, not surprisingly, we find large institutional presence at banks' "guaranteed close", as large-value trades proliferate, although smaller-value trades also exist.[19] Overnight beta is individual stock's overnight return's loading on overnight market return, estimated quarterly using CAPM regression. Spread is time-weighted percent quoted spread during market hours, used as a placebo variable to validate the NYSE fee cut as an exogenous shock.

Finally, daily variables are all aggregated to monthly by taking averages. Given the presence of salient outliers in daily data, we winsorize the overnight returns at 5% tails

[18]We thank the authors for making their code publicly available as supplementary data to the Review of Financial Studies.

[19]Figure C.1 in the appendix plots the distribution of sizes of the off-exchange MOC trades for the stock AAPL. Both extremely large orders and smaller trades exist, and large orders are frequent.

before taking the monthly averages.[20] We further winsorize monthly observations of all the variables at 2% tails.

Price informativeness measure. Price informativeness is usually measured by how well a price tracks the fundamental value of an asset. Close price is, however, unique. Being the last price of the day, it aggregates all the information of the day and is generated by auctions with substantial liquidity and turnovers. Prices right before the close are not good measures of the fundamental value since they do not incorporate all the information of the day, and in some cases are prone to manipulation given the lower liquidity. Neither are prices generated by after hours trades since these trades are infrequent and slim. We recognize the next day's open price as the fundamental value of a stock, to which we compare the close price. In our model and most theoretical literature, fundamental value of the stock is realized in a future period at which investors can liquidate the stock. Aligning with this, for close price, a reasonable measure for the fundamental value it's reflecting would be the next day's open price.[21]

Therefore, our (inverse) measure of the informativeness of the close price is the mean squared error (MSE) of the close price relative to the next day's open price:

$$\text{MSE} = \mathbb{E}\left(\frac{p_{t+1}^{\text{open}} - p_t^{\text{close}}}{p_t^{\text{close}}}\right)^2 \tag{3.1}$$

where p_t^{close} is the close price of day t, p_{t+1}^{open} is the open price of day $t+1$. A lower MSE corresponds to better price informativeness.

[20]Our results are qualitatively similar if we winsorize at, for example, 2% tails, but noisier.

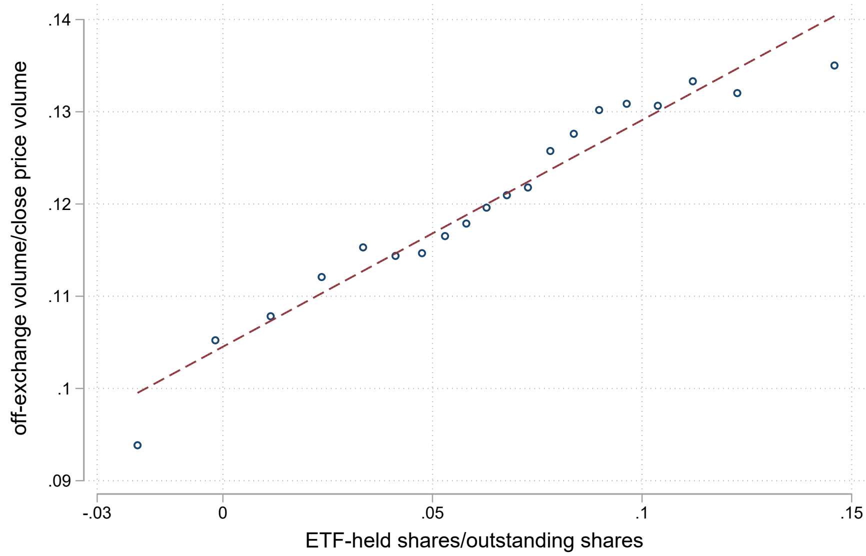
[21]Indeed, next day's open price, beyond the existing information at the market close, also incorporates overnight information, hence is not a perfect measure for the fundamental value *at* the market close. Nevertheless, our results are robust to excluding data within 1 day from earnings announcements days, where overnight information is strong and potentially causes significant changes in the fundamental value.

Note that we scale the difference by the close price itself. This scaling takes into account that prices at higher levels mechanically vary more, and transforms the "closeness" to the fundamental value into return space. Our results are robust to the scaling factor, where we calculate the MSE's using volume-weighted price in the last 5 minutes (p_t^{last5m}), last 15 minutes ($p_t^{last15m}$) or first 5 minutes (p_{t+1}^{open5m}), as the scaling factor.

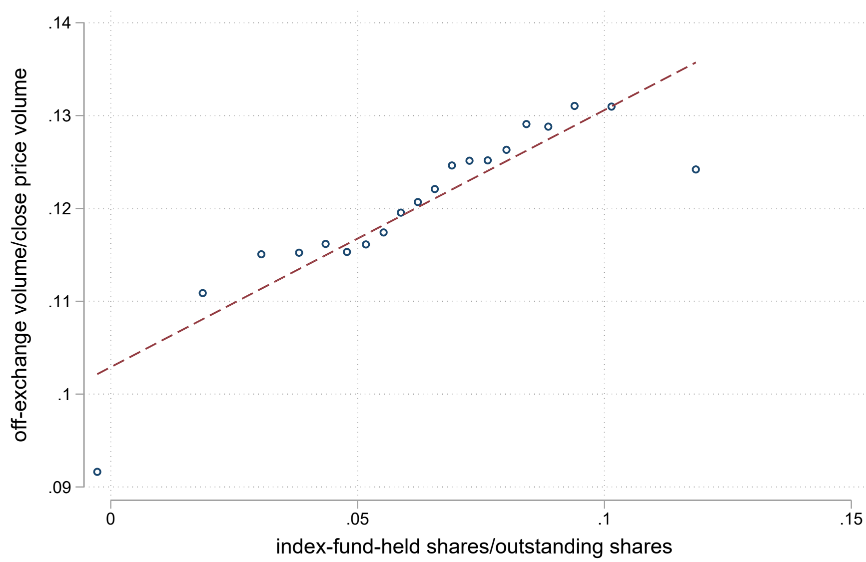
Descriptive findings. Who trade at the banks' "guaranteed close"? In line with anecdotal evidence, we show that the off-exchange MOC volume of a stock is closely related to passive ETF/index fund ownership in the cross-section. Figure 3.2 shows the bin-scatter plots of off-exchange MOC volume as a fraction of the close price volume against ETF ownership, and against index fund ownership. The binscatter plots are based on OLS regression, controlling for a battery of confounding factors including time fixed effects, log(market cap), log(trade volume), volatility, log(total retail volume), log(total volume of trades $\geq$ \$20K in value), log(total volume of trades $\geq$ \$50K in value), after-close volume/total volume, close auction volume/total volume, and overnight beta. The results are robust to not including those controls, but noisier. The bin-scatter plots suggest that as ETF ownership or index fund ownership increases by 1 percent, off-exchange MOC volume as a fraction of the close price volume increases by 0.3 percent.

Figure 3.2. Cross-sectional evidence: relationship between ETF/Index-fund ownership and off-exchange MOC volume

(a) ETF ownership vs. Off-exchange MOC volume/close price volume



(b) Index fund ownership vs. Off-exchange MOC volume/close price volume



Notes. Panel (a) shows the bin-scatter plot between off-exchange MOC volume as a fraction of total volume traded at close price (the "off-exchange fraction"), and ETF ownership of stocks. Panel (b) shows the bin-scatter plot between the "off-exchange fraction" , and index fund ownership of stocks. The sample consists of 4,091 stocks (common shares), from 2015m1 to 2019m12. For each stock, we aggregate the trade volumes from daily data to monthly observations, and then calculate the fractions. ETF ownership and index fund ownership are the average ownerships in 2016. In both panels, we controlled for time fixed effects, log(market cap), log(trade volume), volatility, log(total retail volume), log(total volume of trades $\geq$ \$20K in value), log(total volume of trades $\geq$ \$50K in value), after-close volume/total volume, close auction volume/total volume, and overnight beta in the regressions.

3.3.3. Difference-in-differences: The effects of NYSE fee cut

We use a standard difference-in-differences strategy to estimate the effect of the NYSE close auction fee cut in January 2018 on the MSE of close prices. As discussed in Section 3.3.1, stocks ex-ante more heavily traded at "guaranteed close", and those with higher ETF/index fund ownership are most exposed to the fee cut.

We first drop the NYSE stocks that participated as treated stocks in the 2016 Tick Size Pilot Program. The Pilot increased the tick size for select small stocks, which may force the close price to deviate from the fundamental price. The Pilot ends on September 28, 2018, when the tick size requirements are repealed. This may interfere with the NYSE fee cut quasi-experiment.[22] We rank the remaining stocks by the average fraction of close-price volume (i.e., volume traded at the close price) executed off exchange in 2017. The treated group consists of stocks that rank at top 50% – they are more exposed to the NYSE fee cut. The control group consists of stocks that rank at the bottom 50%. Treated-group stocks have on average a market cap of $22 billion, and ETF/index fund ownership of 8.7% and 8.4%, respectively. Control-group stocks have on average a market cap of $7.1 billion, and ETF/index fund ownership of 4.8% and 3.8%, respectively.

To validate that the fee cut induces large drop in off-exchange MOC volumes for treated stocks (i.e., the first stage), we run a regression of the following form and plot the coefficients β_k's.

$$\frac{\text{off-ex MOC volume}_{i,t}}{\text{close price volume}_{i,t}} = \alpha_i + \lambda_t + \sum_k \beta_k \text{Treat}_i \cdot \mathbb{I}_{t=2017m11+k} + \Gamma X_{i,t} + \epsilon_{i,t}. \tag{3.2}$$

[22]In fact, our results are robust to including those stocks.

where Treat_i is a dummy variable that takes the value one if stock i is in the treated group, and takes the value zero otherwise. $\mathbb{I}_{t=2017m11+k}$ is a dummy variable that takes the value one if time t is k months after 2017m11, and takes the value zero otherwise. Stock fixed effects, time fixed effects, log(market cap), log(total trade volume), volatility, and other control variables, including log(total retail volume), log(total volume of trades $\geq$ \$20K in value), log(total volume of trades $\geq$ \$50K in value), after-close volume/total volume, close auction volume/total volume, and overnight beta are controlled in this regression.

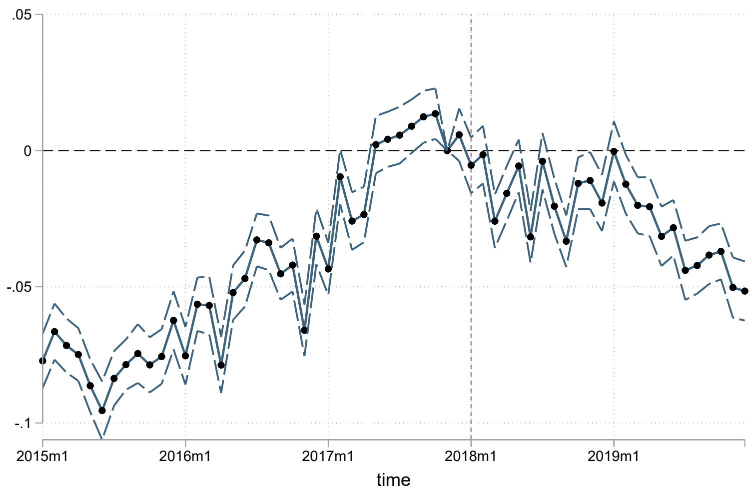
Indeed, it is salient in panel (a) of Figure 3.3, off-exchange MOC volume increased much more rapidly in the treated group prior to the NYSE close auction fee cut. But almost immediately after the fee cut, the off-exchange MOC volume of the treated group stopped growing and started declining, relative to the control group.

Now we study the effect on the informativness of the close price. The major difficulty in difference-in-differences analyses involves separating out pre-existing trends from the dynamic effects of a policy shock. To avoid confounding the two, we first test for pre-existing trends in the MSE measure of price informativeness. Specifically, we run a regression of the same form as Equation (3.2), but replacing the outcome variable with the MSE measure of price informativeness.

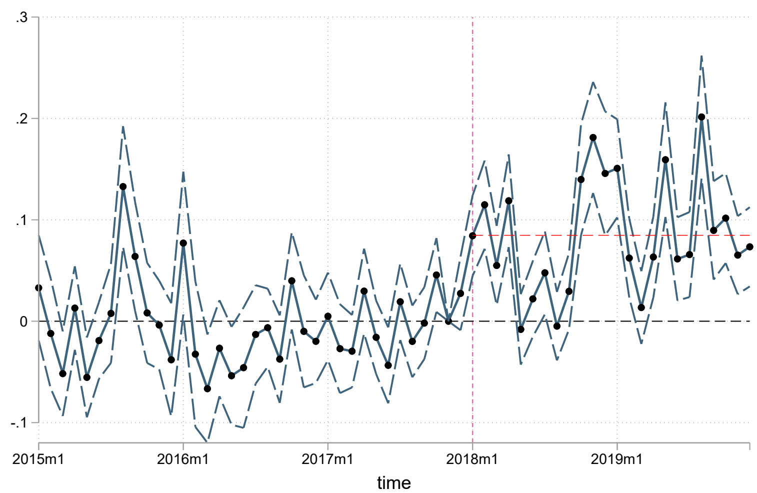
Panel (b) of Figure 3.3 plots the coefficients β_k and the corresponding confidence intervals. We have three findings. First, the MSEs of the treated and control group moved almost perfectly in tandem from 20 months before the NYSE fee cut, so the parallel trend assumption appears to hold well. Second, from September 2015 to March 2016, there appeared to be a drop in MSE for the treated stocks. This same time period was accompanied by the start of the ramp-up of off-exchange MOC volume of treated stocks relative to control stocks. This seems consistent with that off-exchange MOC activity improves close price informativeness, although a causal interpretation is unwarranted since low MSE stocks may select into the treated group. Third, the MSE of treated stocks increased substantially and remained differentially higher than the control stocks after the fee cut.

Figure 3.3. Trends and dynamic responses of off-exchange MOC volume and close price informativeness

(a) First stage: off-exchange MOC volume/close price volume



(b) MSE of close price



Notes. We rank NYSE stocks by the average fraction of close price volume (i.e., volume traded at the close price) executed off exchange in 2017. The treated group consists of stocks that rank at top 50% they are more exposed to the NYSE fee cut. The control group consists of the remaining stocks. Panel (a) shows the treatment effect of the NYSE fee cut on the fraction of close price volume executed off exchange. Panel (b) plots treatment effect of the NYSE fee cut on price informativeness, that is the DID coefficients β_k and 95% confidence intervals estimated from the model: $\text{MSE}_{i,t} = \alpha_i + \lambda_t + \sum_k \beta_k \text{Treat}_i \cdot \mathbb{I}_{t=2017m11+k} + \Gamma X_{i,t} + \epsilon_{i,t}$. We controlled for stock fixed effects, time fixed effects, log(market cap), log(trade volume), volatility, log(total retail volume), log(total volume of trades $\geq$ \$20K in value), log(total volume of trades $\geq$ \$50K in value), after-close volume/total volume, close auction volume/total volume, and overnight beta. Standard errors are clustered at the stock level. The horizontal red dashed line shows the average post-treatment effect, that is, the average of the β_k's.

Finally, we run the following standard difference-in-difference regression.

$$\text{MSE}_{i,t} = \alpha_i + \lambda_t + \beta \text{Treat}_i \cdot \mathbb{I}_{t \geq 2018m1} + \Gamma X_{i,t} + \epsilon_{i,t} \tag{3.3}$$

where $\mathbb{I}_{t \geq 2018m1}$ is a dummy variable that takes the value one if time t is after January 2018, and all the other variables are as defined before in Equation (3.2).

The coefficient of interest is β, which measures the differential change in MSE for the treated stocks and control stocks, holding constant stock-level time-varying characteristics, as well as stock and time fixed effects. Besides MSE (close price as the scaling factor), we also used as a dependent variable the MSEs using volume-weighted price in the last 5 minutes (p_t^{last5m}), last 15 minutes ($p_t^{last15m}$) or first 5 minutes (p_{t+1}^{open5m}) as the scaling factor, as well as the median of $(\frac{p_{t+1}^{open} - p_t^{close}}{p_t^{close}})^2$ (median SE). We also adopt two placebo variables as dependent variables. They are intraday volatility during market hours, and quoted bid-ask spread during market hours. Both of them are important measures of market conditions for a stock, but given that they are measured during market hours, they should not be affected by the NYSE fee cut. To account for serial correlation and stock-specific random shocks, we cluster standard errors at the stock level in all specifications.

Table 3.2 shows the difference-in-difference regression results. We see that the fee cut increased significantly the MSE measures and the median SE of treated stocks, meaning the informativeness of treated group stocks relative to the control group have been worsened by the shock. Not surprisingly, the fee cut did not seem to affect the intraday volatility and quoted spread of the treated stocks.

Table 3.2. DID estimated effects of NYSE fee cut on price informativeness: Main results

	(1) MSE	(2) MSE1	(3) MSE2	(4) MSE3	(5) Median SE	(6) MAD	(7) Volatility	(8) Spread
Treat × Post	0.088***	0.122***	0.084***	0.085***	0.044***	0.052***	0.017	-0.306
	(10.738)	(10.880)	(10.990)	(10.976)	(9.648)	(13.271)	(0.413)	(-1.954)
Volatility	0.051***	0.116***	0.047***	0.047***	0.018***	0.024***		3.812***
	(16.642)	(24.107)	(16.595)	(16.438)	(11.788)	(18.988)		(26.790)
log(Total Volume)	0.144***	0.215***	0.129***	0.132***	0.056***	0.060***	-1.659***	-3.823***
	(11.016)	(11.774)	(10.644)	(10.698)	(7.836)	(10.525)	(-12.615)	(-12.707)
N	66580	66579	66580	66580	66580	66580	66580	66580
Adj. R^2	0.710	0.700	0.723	0.719	0.649	0.782	0.767	0.918
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. The table reports the difference-in-differences regressions estimating the effect of NYSE fee cut on close price informativeness. The sample consists of 1,217 NYSE stocks, and spans from 2015m1 to 2019m12. The dependent variables in columns 1-4, MSE, MSE1, MSE2 and MSE3, are mean squared error measures of price informativeness, calculated from daily data. Specifically, they are the monthly average of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open5m}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close5m}})^2$, and of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close15m}})^2$, respectively. The dependent variable is the monthly median of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$ in column 5, the monthly average of $|\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}}|$ in column 6, volatility during market hours in column 7, and time-weighted percent quoted spread during market hours in column 8. Treat is a dummy that takes the value of one if a stock ranks in top 50% in the average fraction of close price volume (i.e., volume traded at the close price) executed off exchange in 2017 . Post is a dummy that takes the value of one if the time is after the NYSE fee cut time (Jan 2018). The control variables are volatility, log(market cap), log(total volume), log(total retail volume), log(total volume of trades $\geq$ 20K in value), log(total volume of trades $\geq$ 50K in value), after-close volume/total volume, close auction volume/total volume, overnight beta, stock fixed effects and time fixed effects. See definitions of these control variables in the text. Standard errors are clustered at the stock level.

3.3.4. Robustness checks

We take several additional steps to ensure the validity of our research design and the robustness of our estimates.

Alternative treatment designation. One potential concern with the difference-in-difference results is about the designation of the treatment group. We argued that stocks more heavily traded ex-ante at the bank's venue would be more exposed to the fee cut, since they have higher ETF/index fund ownership, while ETF/index funds are more sensitive to fees and in the meantime received larger NYSE fee cuts. We now verify this idea by defining the treatment group with ownership by ETFs and index funds. Specifically, we classify stocks ranked in the top 50% in the sum of ETF and index-fund ownership into the treated group, and the remaining stocks into the control group. Table 3.3 gives the estimation results which are all consistent with the main results.[23]

[23] Quoted spreads of treated stocks are estimated to increased by the fee cut, but not very statistically significant.

Table 3.3. DID estimated effects of NYSE fee cut on price informativeness: Designating treatment group by passive ownership

	(1) MSE	(2) MSE1	(3) MSE2	(4) MSE3	(5) Median SE	(6) MAD	(7) Volatility	(8) Spread
Treat × Post	0.080***	0.107***	0.076***	0.077***	0.036***	0.045***	-0.047	0.288
	(9.868)	(9.643)	(10.061)	(10.048)	(8.056)	(11.590)	(-1.189)	(1.889)
Volatility	0.051***	0.117***	0.047***	0.048***	0.018***	0.024***		3.813***
	(16.564)	(24.014)	(16.512)	(16.355)	(11.780)	(18.752)		(26.806)
log(Total Volume)	0.143***	0.213***	0.128***	0.131***	0.055***	0.059***	-1.670***	-3.718***
	(10.870)	(11.594)	(10.484)	(10.540)	(7.625)	(10.172)	(-12.595)	(-12.328)
N	66569	66568	66569	66569	66569	66569	66569	66569
Adj. R^2	0.710	0.700	0.722	0.719	0.649	0.782	0.767	0.918
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. The table reports the difference-in-differences regressions estimating the effect of NYSE fee cut on close price informativeness. The sample consists of 1,217 NYSE stocks, and spans from 2015m1 to 2019m12. The dependent variables in columns 1-4, MSE, MSE1, MSE2 and MSE3, are mean squared error measures of price informativeness, calculated from daily data. Specifically, they are the monthly average of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open5m}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close5m}})^2$, and of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close15m}})^2$, respectively. The dependent variable is the monthly median of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$ in column 5, the monthly average of $|\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}}|$ in column 6, volatility during market hours in column 7, and time-weighted percent quoted spread during market hours in column 8. Treat is a dummy that takes the value of one if a stock ranks in top 50% in the average fraction of shares held by ETFs and index funds in 2017. Post is a dummy that takes the value of one if the time is after the NYSE fee cut time (Jan 2018). The control variables are volatility, log(market cap), log(total volume), log(total retail volume), log(total volume of trades $\geq$ 20K in value), log(total volume of trades $\geq$ 50K in value), after-close volume/total volume, close auction volume/total volume, overnight beta, stock fixed effects and time fixed effects. See definitions of these control variables in the text. Standard errors are clustered at the stock level.

Excluding earnings announcement days. We take the next day's open price as the fundamental value to which we compare the close price. Next day's open price, beyond the existing information at the market close, also incorporates overnight information, hence is not the fundamental value *at* the market close. For most stocks, earnings announcements are the major overnight information that affects an individual stock's price. One potential concern is that treated stocks may experience stronger earnings news after 2018. To address this, we exclude the data within 1 day, that is, the [-1,0,1] days, from earnings announcements days, before conducting the difference-in-difference estimation. As shown in Table 3.4, our results are robust to excluding these dates.

Table 3.4. DID estimated effects of NYSE fee cut on price informativeness: Excluding earnings announcement days

	(1) MSE	(2) MSE1	(3) MSE2	(4) MSE3	(5) Median SE	(6) MAD	(7) Volatility	(8) Spread
Treat × Post	0.068***	0.100***	0.066***	0.067***	0.042***	0.047***	0.019	-0.313*
	(11.076)	(11.544)	(11.114)	(11.158)	(9.347)	(13.125)	(0.476)	(-1.994)
Volatility	0.035***	0.085***	0.034***	0.034***	0.017***	0.020***		3.870***
	(16.953)	(23.477)	(16.705)	(16.618)	(11.791)	(18.654)		(26.970)
log(Total Volume)	0.085***	0.146***	0.081***	0.081***	0.054***	0.044***	-1.602***	-3.753***
	(9.246)	(10.967)	(8.981)	(8.999)	(7.581)	(8.943)	(-12.631)	(-12.601)
N	66530	66529	66530	66530	66530	66530	66530	66530
Adj. R^2	0.759	0.733	0.759	0.759	0.647	0.794	0.764	0.917
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. The table reports the robustness of the main difference-in-differences regressions estimates to excluding earnings announcement days from the sample. The sample consists of 1,217 NYSE stocks, and spans from 2015m1 to 2019m12. The dependent variables in columns 1-4, MSE, MSE1, MSE2 and MSE3, are mean squared error measures of price informativeness, calculated from daily data, excluding data within 1 day, that is, the [-1,0,1] days, from earnings announcement days. Specifically, they are the monthly average of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open5m}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close5m}})^2$, and of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close15m}})^2$, respectively. The dependent variable is the monthly median of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$ in column 5, the monthly average of $|\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}}|$ in column 6, volatility during market hours in column 7, and time-weighted percent quoted spread during market hours in column 8. Treat is a dummy that takes the value of one if a stock ranks in top 50% in the average fraction of close price volume (i.e., volume traded at the close price) executed off exchange in 2017 . Post is a dummy that takes the value of one if the time is after the NYSE fee cut time (Jan 2018). The control variables are volatility, log(market cap), log(total volume), log(total retail volume), log(total volume of trades $\geq$ 20K in value), log(total volume of trades $\geq$ 50K in value), after-close volume/total volume, close auction volume/total volume, overnight beta, stock fixed effects and time fixed effects. See definitions of these control variables in the text. Standard errors are clustered at the stock level.

Matching specification. Another concern regarding our estimates is that stocks in the treated and control group might be very different in size and many other characteristics, although they are already all NYSE stocks. To alleviate this concern, we conduct a matched sample approach. We match the stocks based on pre-treatment values of log(market cap), log(trading volume), and intraday volatility. For each stock, the closest matching control stock is chosen (with replacement) according to the Mahalanobis distance of the three variables, to constitute the matched control group. Table 3.5 shows that the matched sample yields quantitatively similar results compared to the nonmatched sample. Table 3.6 shows a set of balance test results for the nonmatched samples and the matched samples.

Table 3.5. DID estimated effects of NYSE fee cut on price informativeness: Matching specification

	(1) MSE	(2) MSE1	(3) MSE2	(4) MSE3	(5) Median SE	(6) MAD	(7) Volatility	(8) Spread
Treat × Post	0.068***	0.102***	0.065***	0.066***	0.032***	0.037***	0.028	0.163
	(7.555)	(8.693)	(7.767)	(7.737)	(6.194)	(8.940)	(1.811)	(1.512)
Volatility	0.078***	0.157***	0.071***	0.073***	0.023***	0.035***		4.972***
	(6.942)	(7.738)	(6.935)	(6.937)	(5.304)	(6.933)		(11.969)
log(Total Volume)	0.197***	0.291***	0.181***	0.185***	0.082***	0.089***	-0.505***	-3.625***
	(13.106)	(14.026)	(12.944)	(12.988)	(9.686)	(12.938)	(-7.367)	(-14.842)
N	68186	68186	68186	68186	68186	68186	68186	68186
Adj. R^2	0.734	0.719	0.746	0.743	0.671	0.799	0.620	0.846
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. The table reports the robustness of the main difference-in-differences estimates to using matched treated and control stocks. We perform the matching on market cap, total trade volume, volatility, and overnight beta. The sample consists of 1,217 NYSE stocks, and spans from 2015m1 to 2019m12. The dependent variables in columns 1-4, MSE, MSE1, MSE2 and MSE3, are mean squared error measures of price informativeness, calculated from daily data, excluding data within 1 day, that is, the [-1,0,1] days, from earnings announcement days. Specifically, they are the monthly average of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open5m}-p_t^{close}}{p_t^{close}})^2$, of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close5m}})^2$, and of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close15m}})^2$, respectively. The dependent variable is the monthly median of $(\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}})^2$ in column 5, the monthly average of $|\frac{p_{t+1}^{open}-p_t^{close}}{p_t^{close}}|$ in column 6, volatility during market hours in column 7, and time-weighted percent quoted spread during market hours in column 8. Treat is a dummy that takes the value of one if a stock ranks in top 50% in the average fraction of close price volume (i.e., volume traded at the close price) executed off exchange in 2017. Post is a dummy that takes the value of one if the time is after the NYSE fee cut time (Jan 2018). The control variables are volatility, log(market cap), log(total volume), log(total retail volume), log(total volume of trades $\geq$ 20K in value), log(total volume of trades $\geq$ 50K in value), after-close volume/total volume, close auction volume/total volume, overnight beta, stock fixed effects and time fixed effects. See definitions of these control variables in the text. Standard errors are clustered at the stock level.

Table 3.6. Distribution of variables and balance of matching

	Treated			Control			Comparison	
A. Pre-Matching								
	Mean	Variance	Skewness	Mean	Variance	Skewness	Std-diff	Var-ratio
log(Market Cap)	2.327	1.480	0.225	0.428	2.574	0.674	1.334	0.575
log(Total Volume)	13.933	1.250	-0.051	12.576	2.294	0.104	1.019	0.545
Volatility	0.154	0.181	13.679	1.094	5.894	4.005	-0.539	0.031
Overnight Beta	1.015	0.262	0.280	0.882	0.376	0.254	0.235	0.697
B. Post-Matching (Mahalanobis distance with replacement)								
	Mean	Variance	Skewness	Mean	Variance	Skewness	Std-diff	Var-ratio
log(Market Cap)	2.327	1.480	0.225	2.222	1.360	0.049	0.089	1.089
log(Total Volume)	13.933	1.250	-0.051	13.872	1.076	-0.030	0.056	1.162
Volatility	0.154	0.181	13.679	0.150	0.195	13.831	0.009	0.929
Overnight Beta	1.015	0.262	0.280	1.042	0.243	0.240	-0.055	1.078

Notes. This table reports distributional test statistics of four variables (pre-treatment values) we use in matching treated stocks to control stocks: log(market cap), log(total volume), volatility during market hours, and overnight beta. The table also assesses balance between treated and control group in the means (using the standardized difference) and in the variances (using the variance ratio).

3.3.5. Discussion

How does the NYSE fee cut affect the welfare of index fund/ETF investors? We conduct the following back-of-envelop calculation to shed light on this question.

Suppose there is an index fund that invests in the universe of NYSE stocks using a value weighted portfolio. For simplicity, let us assume the index fund always submit all the orders to the NYSE close auction both before and after the NYSE fee cut. If we define $\mathbb{P}(reversal)$ as the probability that a stock's overnight movement is against the index fund's trade, that is, the stock price goes up yet the index fund sold it at market close, and that the stock price goes down yet the index fund bought it at market close.

Let us denote the overnight return as r, then the overnight profit of the index fund is

$$-\mathbb{E}(|r||\text{reversal})\mathbb{P}(\text{reversal}) + \mathbb{E}(|r||\text{no reversal})\mathbb{P}(\text{no reversal}) \tag{3.4}$$

Empirically we define a stock has an overnight reversal for the index fund if the return from last 5min to close and the return from close to next day's open have opposite sign. The distribution of overnight return r conditional on reversal is empirically similar to the distribution conditional on no reversal. Hence we assume $\mathbb{E}(|r||\text{reversal}) = \mathbb{E}(|r||\text{no reversal})$, and the overnight profit is $[1 - 2\mathbb{P}(\text{reversal})]\mathbb{E}(|r|)$.

For each dollar of transaction, the effect of NYSE fee cut on the index fund's profit is

$$[1 - 2\mathbb{P}(\text{reversal})]\Delta\mathbb{E}(|r|) + \text{fee cut} \times \frac{\text{Total shares}}{\text{Total MV}} \tag{3.5}$$

where we estimate $\Delta\mathbb{E}(|r|)$ to be close to 5 bps, based on regression coefficient in Column 6 of Table 3.2. For the universe of NYSE stocks, our calculation suggests that $\frac{\text{Total shares}}{\text{Total MV}} = 0.18$. The relative fee cut between treated and control group is (3bps - 1bps) = 2bps. As long as $\mathbb{P}(reversal) > 53.6\%$, the welfare of the index fund declines. That is, the benefit of fee cut is outweighed by the cost of the increasing liquidity pressure in the close price.

In our data, $\mathbb{P}(\text{reversal})$ is close to 60%. In fact, the deviation between the last bid-ask midpoint in regular session trading and the close price may better reflect the index fund's trading direction. Bogousslavsky and Muravyev (2021) find that the deviation between close price and last bid-ask midpoint almost always reverses overnight. In that case, $\mathbb{P}(\text{reversal}) = 1$.

3.4. A Model of Dual Trading and Price Discovery at Market Close

In this section, we present a model to better understand the empirical findings. The model is closely tailored to our empirical setting, where traders explicitly choose whether to trade on an exchange, or with a bank, both offering guaranteed execution at the close price, which is set by the market maker on the exchange. We find that the informativeness of a stock's close price is improved when the bank trades based on total orders received, and is increasing in the total volume traded with the bank under mild conditions on parameters.

This result is driven by the bank's trading activity. As long as there is a difference between the two venues in the ratio of informed orders to uninformed orders, the bank can trade profitably solely by observing the net orders it received. In particular, it trades in the same(opposite) direction as the net orders received when the orders contain a higher(lower) proportion of informed orders than that of orders traded on the exchange. While informed traders are worse off due to the increase in price informativeness, the welfare changes of uninformed traders are more nuanced and discussed in details. In order to provide closed-form solutions and sharpen its predictions, our model adopts several simplifying assumptions. We end the section with a thorough discussion of these assumptions, and explain how our model can shed light on other scenarios such as the effect of Robinhood-like e-brokers on price discovery.

3.4.1. Model Setup

There are three periods, denoted by $t = 0, 1, 2$. A single asset has an uncertain liquidation value $v \sim N(0, \sigma^2)$, which is realized and publicly revealed at $t = 2$.

There are m uninformed traders and n informed traders. Each trader buys or sells one unit of the asset when participating.[24] A trader can submit her order either to a market maker or a bank, both executing her order at the same close price p announced at $t = 2$ by the market maker. The market maker and the bank are considered as competitive sectors and break-even in expectation. The market maker determines the close price p. While executing ordering for traders, the bank also submits orders to the market maker itself. All market participants are risk neutral and use market orders only.

A trader is charged a transaction fee φ_m when trading with the market maker and φ_b when trading with the bank per unit of asset traded. φ_m is exogenous, representing the fee charged by the exchange that changes infrequently in practice. φ_b is determined by the bank's break-even condition.[25] Fees are announced at the beginning of $t = 0$. There are convenience costs for traders to trade with the bank, which represent unmodeled factors that deter investors from trading with the banks such as the cost of building connections and contracting with the bank.[26] Convenience costs are heterogeneous among traders, following cumulative distribution functions $G_u : [0, \infty) \to [0, 1]$ and $G_x : [0, \infty) \to [0, 1]$ for

[24] The assumption of unit demand can reflect capital constraints that limit a trader's maximal trade size. The assumption of unit demand is frequently observed in models involving venue choice, for example, Zhu (2014); Hendershott and Mendelson (2000).

[25] These assumptions align with the institutional details in practice, where many banks are competing in the "guaranteed close" business while the primary exchanges are monopoly providers of the close auctions for stocks on each exchange. This does not conflict our assumption that the market maker break-even, since the fee is charged by the exchange instead of the market makers who compete to provide liquidity in the close auction.

[26] In practice, investors can use the "guaranteed close" service only on brokerage accounts at the banks that provide such service. Any incentives that motivate them to choose other brokers over these banks are considered as the convenience costs here. Brokers offer a variety of services to investors by providing efficient execution, market research, and order flow information (Di Maggio et al., 2021). Services are heterogeneous among brokers and investors differ in their demand for these services, so they choose different brokers and are more or less willing to move their portfolios to an account served by these banks to get access to the "guaranteed close business".

liquidity traders and informed traders respectively. Traders make their trading decisions after observing the fees.

The figure below shows the timeline of the model. At the beginning of $t = 0$, the bank announces fee φ_b. Uninformed traders receive liquidity shocks such that they have to buy or sell one unit of the asset.[27] Assume that each uninformed trader has equal probabilities $\frac{1}{2}$ to be a buyer or a seller and liquidity shocks are independent among traders. Then the net liquidity orders u approximately follows a normal distribution $N(0, \sigma_u^2)$ where $\sigma_u^2 = m$ when m is large enough. [28] Random variables v, u are independent and whether a liquidity trader receives a positive or negative shock is independent of her convenience cost for trading via the bank. Each informed trader receives a signal about the value of v that is uniformly distributed $s_i \sim U(v - \sigma_s, v + \sigma_s)$.

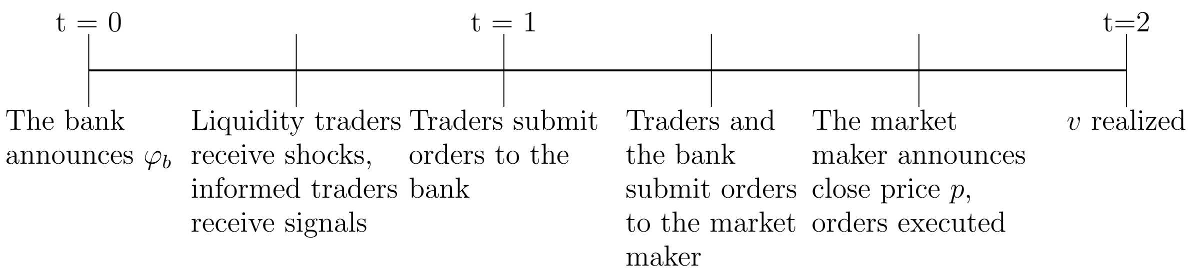


Figure: Model Time Line

At the beginning of $t = 1$, traders choose where to trade and submit their orders. After receiving orders from the traders, the bank can take its own position to trade

[27] We suppose it is very costly for an uninformed trader not to meet her liquidity need, so that she always trades when receiving a shock. The only decision left to be made is the venue choice.

[28] The random number of buyers can be written as $B(m, 1/2)$, where $B(,)$ denotes a binomial distribution. When m is large enough, it can be approximated by a normal distribution $N(m/2, m/4)$. Then the net total orders of these traders $B(m, 1/2) - (m - B(m, 1/2))$ can be approximated by $N(0, m)$. A common rule for this approximation to be appropriate is that everything within 3 standard deviations of its mean is within the range of possible values, which is satisfied in our case as long as $m > 9$.

along or against the net orders received. The bank cannot observe whether an order is informed or not, but knows the composition of its clients given the distributions of traders' features. The bank matches buy and sell orders, including its own position, and submits the remaining order imbalance to the market maker.

At the end of $t = 1$, the close trading ends. The market maker does not observe whether an order is from an informed trader, a liquidity trader or the bank. After collecting all the orders from traders and the bank, the competitive market maker announces the close price p, which equals the expected asset value conditional on the net orders he receives. All orders are then executed at this close price p. Then at $t = 2$, the asset's value v is realized and all participants receive their payoffs.

3.4.2. Equilibrium

An equilibrium consists of the quoting strategy of the market maker, the fee charged by the bank, and the trading strategies of the traders and the bank. In equilibrium, the market maker breaks even in expectation, setting the price p that equals his expected asset value. The bank maximizes its trading profit and breaks even in total profit in expectation. Traders maximize their expected profits. We solve for the equilibrium backwards along the timeline, starting from trading strategies at $t = 1$ given the bank's fee φ_b announced at $t = 0$.

Traders' venue choices. We first characterize the venue choices of the traders. A trader's expected payoff from trading $q_i = \pm 1$ unit of the asset with the market maker is

$$\mathbb{E}_i[v - p]q_i - \varphi_m \tag{3.6}$$

Likewise, the trader's expected payoff from trading via the bank is

$$\mathbb{E}_i[v - p]q_i - (\varphi_b + \gamma_i) \tag{3.7}$$

where $\mathbb{E}_i[\cdot]$ denotes the expectation of trader i, γ_i denotes the convenience cost for trader i to trade with the bank.

Traders choose the trading venue with higher expected payoff, or equivalently, lower total trading cost. Hence a trader trades via bank if

$$\gamma_i < \varphi_m - \varphi_b \tag{3.8}$$

and trades with the market maker otherwise. Let α be the equilibrium fraction of liquidity traders who choose to trade via the bank. The remaining fraction $1-\alpha$ of liquidity traders send their orders to the market maker. Similarly, let θ be the fraction of informed traders who trade via the bank, and the remaining fraction $1 - \theta$ of informed traders send their orders to the market maker. Then we have $\alpha = G_u(\varphi_m - \varphi_b)$, and $\theta = G_x(\varphi_m - \varphi_b)$.

Let u_b, u_m be the net total orders submitted to the bank and the market maker by the liquidity traders. Since liquidity orders are independent among traders, u_b, u_m are independent and approximately normal, and the variances of the net liquidity orders are linear in the fractions of traders who trade via each venue. Then, since $u \sim N(0, \sigma_u^2)$, we have

$$\begin{aligned} u_b &\sim N(0, \alpha\sigma_u^2) \\ u_m &\sim N\left(0, (1-\alpha)\sigma_u^2\right) \end{aligned} \tag{3.9}$$

An informed trader buys if $\mathbb{E}[v - p|s_i] > \min(\varphi_m, \varphi_b + \gamma_i)$ and sells if $-\mathbb{E}[v - p|s_i] > \min(\varphi_m, \varphi_b + \gamma_i)$. When σ_s is large enough, the total informed orders is approximately $x = \eta v$, where $\eta = \frac{m}{\sigma_s}$.[29] Then given the fractions of informed traders trading via each venue shown above, the net total informed orders submitted to the bank (x_b) and the market maker(x_m) are

$$x_b = \theta \eta v$$

$$x_m = (1 - \theta)\eta v \tag{3.10}$$

Bank's trading strategy. Given the liquidity and informed orders submitted to each trading venue, we now characterize the bank's trading strategy. The net total orders received by the bank is $y = u_b + \theta\eta v$, with variance $\sigma_y^2 = \alpha\sigma_u^2 + \theta^2\eta^2\sigma^2$. When $\sigma_y^2 > 0$, that is, the bank receives a positive measure of orders, it takes its own position $d(y)$ based on its expectations given the net orders received and the distributions of traders' features. $d(y)$ can be either along or against the direction of the net orders the bank receives. The bank's expected trading profit is

$$\mathbb{E}[v - p|y]d(y) \tag{3.11}$$

Here we assume that the bank's position $d(y)$ does not incur changes in transaction costs to preserve tractability of the model. [30]

[29]Let $\pi(s_i) = \mathbb{E}[|v - p||s_i]$ and guess it is strictly increasing in $|s_i|$, which can be verified in the following equilibrium. Then $\pi(s_i)$ is reversible for $s_i \geq 0$. Given σ_s, the total informed orders is $x = \frac{m}{\sigma_s}v$ for $v \in (-\sigma_s + \pi^{-1}(\mu_m), \sigma_s - \pi^{-1}(\mu_m))$. Since v is normal, the tail probabilities converges to zero when σ_s goes to infinity so the approximation is valid.

[30]The incentive to save transaction fees paid to the exchange may make the bank more willing to provide liquidity with its own holdings and trade against the net orders received.

We restrict our attention to rational expectations equilibria in which the price is linear in the net orders received by the market maker. Guess that the market maker applies a linear price setting rule $p = \lambda z$, where $z = x + u + d$ is the net orders received by the market maker. Then we can write the bank's trading problem as

$$\max_d E[v - p|y]d = \left(\left[(1-\lambda\eta)\frac{\theta\eta\sigma^2}{\sigma_y^2} - \lambda\frac{\alpha\sigma_u^2}{\sigma_y^2}\right]y - \lambda d\right)d \tag{3.12}$$

By solving the bank's optimal trading strategy and combining it with the market maker's price setting rule, $p = E[v|z]$, where $z = u + x + d$ is the net orders received by the market maker, we get a linear equilibrium as in the following.

Proposition 7. *If $\sigma_y^2 > 0$, a linear Nash equilibrium of the model described above is given by*

$$p(z) = \lambda z \tag{3.13}$$

$$d(y) = Ky \tag{3.14}$$

where $\lambda = \frac{(1+K\theta)\eta\sigma^2}{((1+K)^2\alpha+(1-\alpha))\sigma_u^2+(1+K\theta)^2\eta^2\sigma^2}$, $K = \frac{-B+\sqrt{B^2-4AC}}{2A}$, *with* $A = \sigma_y^2, B = \theta\eta^2\sigma^2 + (2-\theta)\frac{\alpha}{\theta}\sigma_u^2, C = \left(\frac{\alpha}{\theta} - 1\right)\sigma_u^2$ *when* $\theta > 0$, *and* $K = -\frac{1}{2}$ *when* $\theta = 0, \alpha > 0$.[31]

Proofs are in the Appendix.

In this equilibrium, the trading strategy of the bank and the quoting strategy of the market maker are both linear in the net total orders they receive, and determined only by the exogenous parameters and the bank's fee announced at $t = 0$. Some interesting results about the bank's trading strategy are worthwhile to note here. First, K has the

[31]When $\alpha = \theta = 0$, we get the trivial equilibrium with no bank where $\lambda = \frac{a\sigma^2}{\sigma_u^2+a^2\sigma^2}$ and K is not defined.

same sign as $\theta - \alpha$. When $\theta > \alpha$, that is, the fraction of informed traders trading via the bank exceeds that of the liquidity traders, then $K > 0$ — the bank takes its own position in the same direction as the net orders it receives. When $\theta < \alpha$, that is, the fraction of liquidity traders trading via the bank exceeds that of the informed traders, then $K > 0$ — the bank takes its position in the opposite direction to the net orders it receives. This is because if $\theta > \alpha$, the bank perceives the orders received to be more informative than orders on the whole market, so following them is profitable. If $\theta < \alpha$, the bank perceives the orders received to be more noisy than orders on the whole market, so trading against them is profitable due to the price impact of liquidity traders. Second, when $\theta = 0$, the bank's trading strategy is $K = -\frac{1}{2}$. In this case, the bank only receives liquidity orders, and makes a profit by providing liquidity to them. Third, when $\alpha = 0$, $d = \frac{1}{2}\left(\sqrt{\eta^2 + 4\frac{\sigma_u^2}{\sigma^2}} - \eta\right)v$, the bank's trading position does not depend on θ as long as it is positive. This is because when the bank only receives informed orders, it can perfectly infer the value of the asset from just a small mass of informed orders. Then it behaves in the same way as a large informed trader who perfectly observes the asset value.

Fee setting of the bank. Since the bank is competitive, the fee φ_b announced at $t = 0$ is determined by the bank's break-even condition, given the equilibrium trading strategies of the market participants at $t = 1$.

$$0 = \mathbb{E}\left[\mathbb{E}[v - p|y]d(y)|\varphi_b\right] + \mathbb{E}\left[|u_b^+| + |u_b^-| + |\theta\eta v||\varphi_b\right]\varphi_b - \mathbb{E}\left[|u_b + \theta\eta v||\varphi_b\right]\varphi_m \tag{3.15}$$

where u_b^+, u_b^- are buying and selling orders from uninformed traders received by the bank. It is hard to solve for φ_b in closed form generally. However, in order to get the equilibrium results at $t = 1$ solved above, it is enough to show that the bank does participate with $\varphi_b < \varphi_c$ in equilibrium, which is shown below.

Proposition 8. *If $G_u(0) > 0$, there exists $\varphi_b < \varphi_m$ such that the above break-even condition is satisfied.* [32]

Proofs are in the Appendix.

3.4.3. Price informativeness

The MSE measure of price informativeness. Now we measure the informativeness of the close price p, in terms of how well it reveals the fundamental asset value v. We define that the close price is more informative when it is "closer" to v, and we use the scaled mean square error to measure this closeness.

We calculate the MSE of p and scale it by the variance of the asset value σ^2 to get the scaled MSE

$$\frac{MSE}{\sigma^2} = \frac{\mathbb{E}\left[\mathbb{E}[(v-p)^2|v]\right]}{\sigma^2} = \frac{1}{\xi+1} \tag{3.16}$$

where $\xi = \frac{(1+K\theta)^2\eta^2\sigma^2}{[(1+K)^2\alpha+1-\alpha]\sigma_u^2}$.

Improvement in price informativeness. In order to analyze the effect of the bank's guaranteed close service on close price informativeness, we compare the scaled MSE in the equilibrium where bank conducts the guaranteed close service, with the scaled MSE in the equilibrium without the bank. Observe that the equilibrium without the bank

[32] The condition $G_u(0) > 0$ is sufficient and means that a positive proportion of uninformed traders have zero convenience cost. They can represent existing clients of the bank. We are assuming that m is large enough such that the normal approximation applies at $G_u(0)$.

coincides the equilibrium in which the fractions of traders trading via the bank is 0, i.e., $\alpha = \theta = 0$. So the scaled MSE in the "no bank" equilibrium is

$$\frac{MSE_{\text{no bank}}}{\sigma^2} = \frac{1}{\xi_{\text{no bank}} + 1} \tag{3.17}$$

where $\xi_{\text{no bank}} = \frac{\eta^2\sigma^2}{\sigma_u^2}$. Comparing this to the above equilibrium result with the bank, we get the following result.

Proposition 9. *The variance of informed orders increases more than the variance of liquidity orders due to dual trading of the bank, that is,*

$$(1 + K\theta)^2 \geq (1 + K)^2\alpha + 1 - \alpha \tag{3.18}$$

Having the bank conducting guaranteed close service improves the informativeness of p*.*

Proofs are in the appendix.

The above inequality is strict as long as $\alpha \neq \theta$ and $\alpha > 0$, so that the bank does trade for profit in equilibrium. Such trading activity improves the informativeness of the close price because it increases the ratio of informed orders relative to uninformed orders, measured by their variances, received by the market maker.

Comparative statics. Having shown that having the bank improves the informativeness of the close price, we next discuss how the bank's share of orders received from traders affects the price informativeness. We do this by comparative statics on how the above measure of price informativeness changes with the parameters and get the following result.

Proposition 10. *Let* $\delta = \frac{\theta}{\alpha}$ *be the ratio of the proportion of informed traders trading with the bank relative to the proportion of uninformed traders trading with the bank,* $\zeta =$

$\frac{\alpha\sigma_u^2+\theta\eta^2\sigma^2}{\sigma_u^2+\eta^2\sigma^2}$ *be a proxy of the proportion of traders' trading volume executed by the bank, then we can rewrite* $\xi\left(\alpha(\varphi_m),\theta(\varphi_m)\right)$ *as* $\xi\left(\delta(\varphi_m),\zeta(\varphi_m)\right)$*, and*

$$\frac{\partial\xi}{\partial\varphi_m}=\frac{\partial\xi}{\partial\delta}\frac{\partial\delta}{\partial\varphi_m}+\frac{\partial\xi}{\partial\zeta}\frac{\partial\zeta}{\partial\varphi_m} \tag{3.19}$$

where $\frac{\partial\xi}{\partial\zeta}\geq 0$ *and the inequality is strict when* $\delta\neq 1$*. The effect of a fee change by the exchange on the price informativeness can be decomposed into a ratio effect and a volume effect, where the volume effect, i.e., the effect of the bank's volume share on price informativeness is positive, and the ratio effect is mixed.*

Proofs are in the appendix.

The result here exactly aligns with our empirical findings, which is not surprising. Since uninformed traders are independent with each other, the bank infers more information about the overall orders when it receives more orders and trades more. Since the bank's trading is beneficial for the informativeness of the close price, as we have analyzed above, the increase in such trading improves the price informativeness. For example, when $\theta=0$ and $\alpha>0$, the bank provides liquidity to the liquidity orders it receives by trading against half of the trading demand ($K=-\frac{1}{2}$). In this case, the bank provides more liquidity when it receives more orders, making the price more informative. When there is a fee cut on the exchange, the bank's volume share decreases and the volume effect reduces the price informativeness. To see that the ratio effect is mixed, consider when δ is slight above or below 1. Since the informativeness with $\delta=1$ is equivalent to the no bank result that is lower than informativeness with $\delta\neq 1$, an increase in δ makes it further away from 1 and improves the price informativeness when δ is slightly above 1, while making it closer to 1 and reduces the price informativeness when δ is slightly below 1. Our empirical finding that the fee cut reduces price informativeness can be interpreted as the two effects have the same sign or the volume effect is dominating.

3.4.4. Changes in the profits of the traders

Now we discuss the effect of dual trading by the bank on the profits of the traders, that is, the expected gains of the informed traders and the expected losses of the liquidity traders.

Proposition 11. *The bank's business decreases informed traders' expected gain before fees. It improves market depth and decreases the expected loss before fees of liquidity traders who still directly trade with the exchange if and only if $\theta > \alpha$ and $(\theta - 3\alpha)\sigma_u^2 < 2\theta\eta^2\sigma^2$. It decreases the expected loss before fees of liquidity traders who trade with the bank if and only if $\theta < \alpha$ and $(1-\theta)(1+K\theta)\eta^2\sigma^2 + [1+\theta-2\alpha+(\theta-\alpha)K]\sigma_u^2 > 0$.*

Proofs are in the appendix.

It is not surprising that the expected gain of the informed traders decrease given our previous result that price informativeness improved. If the bank receives a larger fraction of informed orders than uninformed orders, it chooses $K > 0$ and competes with the informed traders, reducing their profits. If the bank receives a smaller fraction of informed orders than uninformed orders, it chooses $K < 0$ and reduces market depth, still hurting the profits of the informed even if they do not trade with the bank.

The welfare effects on the liquidity traders are mixed. For a liquidity trader, the welfare effect largely depends on whether she trades with the bank and whether the bank receives a larger fraction of liquidity orders than informed orders. A liquidity trader benefit from the bank's business if she trades with the bank and it provides liquidity to trades with $K < 0$, or if she does not trade with the bank and it amplifies orders received with $K > 0$. Notice that such difference does not affect the venue choices of the liquidity traders since they are small and price taking. The results are calculated for all liquidity traders who trade with the bank and all who do not, instead of each individual trader, and they are not able to coordinate.

3.4.5. Discussion

Lastly, we discuss the robustness of our results to assumption changes and the generality of our model implications on scenarios beyond trading at the market close.

The venue choices are modeled in a largely exogenous way in our model, driven by a heterogeneous convenience cost drawn from arbitrary cumulative distribution functions G_u and G_x. That allows us to generally analyze any arbitrary sorting of traders between the two venues with $(\alpha, \theta) \in [0, 1] \times [0, 1]$. Therefore, our results apply to any specific setting that models the venue choices endogenously, which can be treated as a special case of our model. The improvement of price informativeness due to the introduction of a bank is strict if and only if $\frac{\theta}{\alpha} \neq 1$, that is, the fraction of informed traders trading with the bank is different from that of uninformed traders. Such condition can be satisfied as long as the incentives driving traders' venue choices are not identical between informed and uninformed traders. This is usually true in models with endogenous venue choices due to price or execution related incentives. For example, if we apply our model to a regular darkpool like in Zhu (2014), execution risks make uninformed traders more likely to trade in the darkpool compared to informed traders.

Our model is built on a stylized setting following the one period model in Kyle (1985) and can be applied generally to trading scenarios with multiple trading venues where the brokers are able to trade, which is quite common in most markets. In order to focus on the role of the bank in our specific empirical setting, we assume in our model that only the bank conducts dual trading and traders can directly trade with the market maker. Such assumption does not drive our results on price informativeness, which still apply if everyone trades via dual trading brokers, as long as the brokers receive different fractions of informed orders relative to uninformed order. In practice, the variety of service offered by brokers can make some of them more attractive to informed traders while others more attractive to uninformed traders.

3.5. Conclusion

In this paper, we formally study the impact of "guaranteed close" on the informativeness of the close price, both empirically and theoretically. Using the NYSE close auction fee cut in January 2018 as a policy shock, we provide quasi-experimental evidence that "guaranteed close" improves price discovery. Our unique empirical setting is ideal for testing the effect of dual trading: banks trade on their own accounts after viewing order flows of customers, and a quasi-experimental shock reduces the order flow to the banks; designated market makers clear the market in a single auction and set the price – these features correspond to the framework that the dual-trading literature usually builds on. Our empirical finding cannot be explained by the predictions of the dual-trading literature.

We build a model and provide a novel mechanism through which dual-trading improves price discovery. In our model, traders explicitly choose between two trading venues (i.e., the bank and the close auction) based on transaction costs. As long as the proportions of informed orders relative to uninformed orders are different between two venues, the bank can infer information from the net total orders and trade profitably. Such trading activity amplifies the proportion of informed orders received by the market maker and improves price informativeness.

The US brokerage industry has been undergoing rapid changes in its landscapes nowadays. While some brokers still charge commission fee and offer explicit execution price, e-brokers like Robinhood flock to charge zero commission fee and sell order flow data to sophisticated investment firms. The former brokers would be comparable to the close auction in our paper, and the latter the bank's "guaranteed close". Our model and empirical evidence shed light on the effect of these changes on price discovery. But of course, the jury is still out when it comes to the impact of these changes on liquidity, fairness, and aggregate investor welfare.

References

Admati, Anat R, Peter M DeMarzo, Martin F Hellwig, and Paul Pfleiderer (2018), "The leverage ratchet effect," *The Journal of Finance*, 73, 145–198.

Ai, Hengjie, Murray Z Frank, and Ali Sanati (2020), "The trade-off theory of corporate capital structure,".

Ang, Andrew, Vineer Bhansali, and Yuhang Xing (2010), "Taxes on tax-exempt bonds," *The Journal of Finance*, 65, 565–601.

Auerbach, Alan J (1979), "Share valuation and corporate equity policy," *Journal of Public Economics*, 11, 291–305.

Auerbach, Alan J (1981), "TAX INTEGRATION AND THE" NEW VIEW" OF THE CORPORATE TAX: A 1980'S PERSPECTIVE," In Proceedings of the Annual Conference on Taxation Held under the Auspices of the National Tax Association-Tax Institute of America, 21–27, JSTOR.

Auerbach, Alan J (2002), "Taxation and corporate financial policy," *Handbook of public economics*, 3, 1251–1292.

Baker, Malcolm and Jeffrey Wurgler (2002), "Market timing and capital structure," *The journal of finance*, 57, 1–32.

Bakke, Tor-Erik and Toni M Whited (2010), "Which firms follow the market? An analysis of corporate investment decisions," *The Review of Financial Studies*, 23, 1941–1980.

Baldauf, Markus, Christoph Frei, and Joshua Mollner (2021), "Principal Trading Arrangements: When Are Common Contracts Optimal?," *Management Science*.

Baltussen, Guido, Sjoerd van Bekkum, and Zhi Da (2019), "Indexing and stock market serial dependence around the world," *Journal of Financial Economics*, 132, 26–48.

Barbon, Andrea, Marco Di Maggio, Francesco Franzoni, and Augustin Landier (2019), "Brokers and order flow leakage: Evidence from fire sales," *The Journal of Finance*, 74, 2707–2749.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi (2018), "Do ETFs increase volatility?," *The Journal of Finance*, 73, 2471–2535.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song (2021), "What do mutual fund investors really care about?," *The Review of Financial Studies*, forthcoming.

Berk, Jonathan and Peter DeMarzo (2011), "Corporate Finance, global ed," *Essex: Person Education Limited*.

Bhattacharya, Nilabhra, Ervin L. Black, Theodore E. Christensen, and Richard D. Mergenthaler (2007), "Who trades on pro forma earnings fnformation?," *The Accounting*

Review, 82, 581–619.

Blouin, Jennifer, John E Core, and Wayne Guay (2010), "Have the tax benefits of debt been overestimated?," *Journal of Financial Economics*, 98, 195–213.

Boehmer, Ekkehart, Charlies Jones, Xiaoyan Zhang, and Xinran Zhang (2021), "Tracking retail investor activity," *The Journal of Finance*, 76, 2249–2305.

Bogousslavsky, Vincent and Dmitriy Muravyev (2021), "Who trades at the close? Implications for price discovery and liquidity," *Available at SSRN.*

Bolton, Patrick, Hui Chen, and Neng Wang (2014), "Debt, taxes, and liquidity," Technical report, National Bureau of Economic Research.

Bolton, Patrick, Neng Wang, and Jinqiang Yang (2021), "Leverage dynamics under costly equity issuance," *NBER Working Paper*, 26802.

Bond, Philip, Alex Edmans, and Itay Goldstein (2012), "The real effects of financial markets," *Annu. Rev. Financ. Econ.*, 4, 339–360.

Brogaard, Jonathan and Jing Pan (2021), "Dark pool trading and information acquisition," *The Review of Financial Studies, Forthcoming.*

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

Chakravarty, Sugato and Kai Li (2003), "An examination of own account trading by dual traders in futures markets," *Journal of Financial Economics*, 69, 375–397.

Chen, Daniel and Darrell Duffie (2021), "Market fragmentation," *American Economic Review*, 111, 2247–74.

Cooley, Thomas F and Vincenzo Quadrini (2001), "Financial markets and firm dynamics," *American economic review*, 91, 1286–1310.

Crouzet, Nicolas and Fabrice Tourre (2021), "Can the cure kill the patient? Corporate credit interventions and debt overhang," *Corporate credit interventions and debt overhang (June 1, 2021).*

Dangl, Thomas and Josef Zechner (2021), "Debt maturity and the dynamics of leverage," *The Review of Financial Studies*, 34, 5796–5840.

Dannhauser, C. D. and Jeffery Pontiff (2019), "Flow," *Working Paper, Boston College.*

Dávila, Eduardo and Benjamin M Hébert (2019), "Optimal Corporate Taxation Under Financial Frictions," Technical report, National Bureau of Economic Research.

Demarzo, Peter M (2019), "Presidential address: Collateral and commitment," *The Journal of Finance*, 74, 1587–1619.

DeMarzo, Peter M and Zhiguo He (2021), "Leverage dynamics without commitment," *The Journal of Finance*, 76, 1195–1250.

Di Maggio, Marco, Mark Egan, and Francesco Franzoni (2021), "The value of intermediation in the stock market," *Journal of Financial Economics.*

Diamond, Douglas W and Zhiguo He (2014), "A theory of debt maturity: the long and short of debt overhang," *The Journal of Finance*, 69, 719–762.

Ernst, Thomas, Jonathan Sokobin, and Chester Spatt (2021), "The value of off-exchange data," *Working Paper.*

Faulkender, Michael and Jason M Smith (2016), "Taxes and leverage at multinational corporations," *Journal of Financial Economics*, 122, 1–20.

Fischer, Edwin O, Robert Heinkel, and Josef Zechner (1989), "Dynamic capital structure choice: Theory and tests," *The journal of finance*, 44, 19–40.

Fishman, Michael J and Francis A Longstaff (1992), "Dual trading in futures markets," *The Journal of Finance*, 47, 643–671.

Fortune, Peter (1996), "Do municipal bond yields forecast tax policy?," *New England Economic Review*, 29–49.

Gamba, Andrea and Alexander Triantis (2008), "The value of financial flexibility," *The journal of finance*, 63, 2263–2296.

Goldstein, Michael A, Paul Irvine, Eugene Kandel, and Zvi Wiener (2009), "Brokerage commissions and institutional trading patterns," *The Review of Financial Studies*, 22, 5175–5212.

Goldstein, Robert, Nengjiu Ju, and Hayne Leland (2001), "An EBIT-based model of dynamic capital structure," *The Journal of Business*, 74, 483–512.

Graham, John R (1996a), "Debt and the marginal tax rate," *Journal of financial Economics*, 41, 41–73.

Graham, John R (1996b), "Proxies for the corporate marginal tax rate," *Journal of financial economics*, 42, 187–221.

Graham, John R (2000), "How big are the tax benefits of debt?," *The journal of finance*, 55, 1901–1941.

Hackbarth, Dirk and David C Mauer (2012), "Optimal priority structure, capital structure, and investment," *The Review of Financial Studies*, 25, 747–796.

He, Zhiguo and Konstantin Milbradt (2014), "Endogenous liquidity and defaultable bonds," *Econometrica*, 82, 1443–1508.

He, Zhiguo and Konstantin Milbradt (2016), "Dynamic debt maturity," *The Review of Financial Studies*, 29, 2677–2736.

He, Zhiguo and Wei Xiong (2012), "Rollover risk and credit risk," *The Journal of Finance*, 67, 391–430.

Heider, Florian and Alexander Ljungqvist (2015), "As certain as debt and taxes: Estimating the tax sensitivity of leverage from state tax changes," *Journal of financial economics*, 118, 684–712.

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

Hennessy, Christopher A and Toni M Whited (2005), "Debt dynamics," *The journal of finance*, 60, 1129–1165.

Hennessy, Christopher A and Toni M Whited (2007), "How costly is external financing? Evidence from a structural estimation," *The Journal of Finance*, 62, 1705–1745.

Hong, Harrison and Jiang Wang (2000), "Trading and returns under periodic market closures," *The Journal of Finance*, 55, 297–354.

Huddart, Steven, John S Hughes, and Carolyn B Levine (2001), "Public disclosure and dissimulation of insider trades," *Econometrica*, 69, 665–681.

Ivanov, Ivan, Luke Pettit, and Toni M Whited (2020), "Taxes depress corporate borrowing: Evidence from private firms," *Available at SSRN 3694869*.

Ju, Nengjiu, Robert Parrino, Allen M Poteshman, and Michael S Weisbach (2005), "Horses and rabbits? Trade-off theory and optimal capital structure," *Journal of Financial and Quantitative analysis*, 40, 259–281.

King, Mervyn A (1974), "Taxation and the cost of capital," *The Review of Economic Studies*, 41, 21–35.

King, Mervyn A (1977), Public policy and the corporation 3, London: Chapman and Hall; New York: Wiley.

Kothari, SP, Travis L Johnson, and Eric C So (2021), "Commission Savings and Execution Quality for Retail Trades," *Available at SSRN*.

Kueng, Lorenz (2018), "Tax news shocks and consumption," *Swiss Finance Institute Research Paper*.

Kurshev, Alexander and Ilya A Strebulaev (2015), "Firm size and capital structure," *Quarterly Journal of Finance*, 5, 1550008.

Kyle, Albert S (1985), "Continuous auctions and insider trading," *Econometrica: Journal of the Econometric Society*, 1315–1335.

Lee, Charles M.C. and Balkrishna Radhakrishna (2000), "Inferring investor behavior: Evidence from TORQ data," *Journal of Financial Markets*, 3, 83–111.

Leland, Hayne et al. (1994), "Bond prices, yield spreads, and optimal capital structure with default risk," Technical report, University of California at Berkeley.

Leland, Hayne E (1994), "Corporate debt value, bond covenants, and optimal capital structure," *The journal of finance*, 49, 1213–1252.

Leland, Hayne E (1998), "Agency costs, risk management, and capital structure," *The Journal of Finance*, 53, 1213–1243.

Leland, Hayne E and Klaus Bjerre Toft (1996), "Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads," *The Journal of Finance*, 51, 987–1019.

Lemmon, Michael L, Michael R Roberts, and Jaime F Zender (2008), "Back to the beginning: persistence and the cross-section of corporate capital structure," *The journal of finance*, 63, 1575–1608.

Lewellen, Jonathan and Katharina Lewellen (2006), "Internal equity, taxes, and capital structure," *Unpublished Working Paper, Dartmouth College*.

Lines, Anton (2022), "Do Institutional Incentives Distort Asset Prices?," *Available at SSRN 2873739*.

Longstaff, Francis A (2011), "Municipal debt and marginal tax rates: Is there a tax premium in asset prices?," *The Journal of Finance*, 66, 721–751.

Miller, Merton H (1977), "Debt and taxes," *the Journal of Finance*, 32, 261–275.

Modigliani, Franco and Merton H Miller (1963), "Corporate income taxes and the cost of capital: a correction," *The American economic review*, 53, 433–443.

Poterba, James M and Lawrence H Summers (1984), "The economic effects of dividend taxation," Technical report, National Bureau of Economic Research.

Röell, Ailsa (1990), "Dual-capacity trading and the quality of the market," *Journal of Financial Intermediation*, 1, 105–124.

Sarkar, Asani (1995), "Dual trading: Winners, losers, and market impact," *Journal of Financial Intermediation*, 4, 77–93.

Schwert, Michael (2017), "Municipal bond liquidity and default risk," *The Journal of Finance*, 72, 1683–1722.

SEC DERA (2017), "Bats market close: Off-exchange closing volume and price discovery," *SEC memorandum*.

Shanthikumar, Devin (2003), "Small and large trades around earnings announcements: Does trading behavior explain post-earnings-announcement drift?," *Working Paper*.

Shevlin, Terry (1990), "Estimating corporate marginal tax rates with asymmetric tax treatment of gains and losses," *Journal of the American Taxation Association*, 11, 51–67.

Strebulaev, Ilya A (2007), "Do tests of capital structure theory mean what they say?," *The journal of finance*, 62, 1747–1787.

Strebulaev, Ilya A and Baozhong Yang (2013), "The mystery of zero-leverage firms," *Journal of Financial Economics*, 109, 1–23.

Titman, Sheridan and Sergey Tsyplakov (2007), "A dynamic model of optimal capital structure," *Review of Finance*, 11, 401–451.

WSJ (2018), "Goldman cashes in on passive-investing boom with big 4 P.M. trade," *The Wall Street Journal (August 26, 2018)*.

Yang, Liyan and Haoxiang Zhu (2020), "Back-running: Seeking and hiding fundamental information in order flows," *The Review of Financial Studies*, 33, 1484–1533.

Ye, Mao (2010), "Non-execution and market share of crossing networks," *Available at SSRN 1719016*.

Zhu, Haoxiang (2014), "Do dark pools harm price discovery?," *The Review of Financial Studies*, 27, 747–789.

APPENDIX A

Appendix for Chapter 1

A.1. Proofs

A.1.1. Proof of proposition 4

I characterize the relations between regions of financing strategies where different equilibrium conditions hold by checking the limiting conditions for extreme values of state variables and the boundary conditions between regions. By continuity, the regions in which conditions 1 and 3 in Proposition 3 hold (equity issuing region and dividend distributing region) do not connect. Therefore, in the followings, I check equilibrium conditions at bankruptcy, when $\frac{Y}{F}$ converges to infinity and at the boundaries where a break-even by debt region connects to an equity issuing region or a dividend distributing region.

When bankruptcy is triggered, both the debt price and the first order partial derivative of the value function are zero, that is, $p^b(Y,F) = V^b_F(Y,F) = 0$, where $p^b(Y,F), V^b(Y,F)$ denote the debt price and the value function at a pair of (Y,F) such that bankruptcy is triggered.

When $\frac{Y}{F} \to \infty$, the firm's securities converge to the risk-free values,

$$p^{rf}(Y,F) = \frac{(1-\tau_b)c + m}{r+m} \tag{A.1}$$

$$V^{rf}(Y,F) = V^{rf}(Y,0) - (1-\tau_e)\frac{(1-\tau_c)c + m}{r+m}F \tag{A.2}$$

where $p^{rf}(Y,F), V^{rf}(Y,F)$ denote the risk-free limit of the debt price and the value function. If $\tau_b > \tau_c$, then $-\frac{V_F^{rf}(Y,F)}{1-\tau_e} > p^{rf}(Y,F) > -V_F^{rf}(Y,F)$, and $\phi^{rf}(Y,F) = \bar{\phi}^{rf}(Y,F)$. The firm spends all free cash on debt repurchase when $\frac{Y}{F}$ is large enough, to save the difference between personal income tax rates priced in bonds and the corporate tax rates. The tax benefits dominate the debt ratchet effect of debt repurchase when the remaining debt is low enough. If $\tau_b < \tau_c$, then $-\frac{V^{rf}(Y,F)}{1-\tau_e} < p^{rf}(Y,F)$ and the equilibrium security prices cannot converge to the risk-free valuations since otherwise the firm will issue debt discretely to take advantage of the high debt price until $-\frac{V^{rf}(Y,F)}{1-\tau_e} = p^{rf}(Y,F)$. That is because investors always expect the firm to lever up when leverage is low.

Then I analyze the boundary conditions between regions. Suppose there exists a region of (Y,F) such that the first order condition $p(Y,F) = -V_F(Y,F)$ holds (equity issuing region). Then at the boundary between this region and the region in which $-\frac{V^{rf}(Y,F)}{1-\tau_e} > p(Y,F) > -V_F(Y,F)$ (break-even by debt region),

$$\begin{aligned} -rV_F(Y,F) =& (1-\tau_c)c - p_F(Y,F)\phi + m(1+V_F(Y,F)) - (\phi - mF)V_{FF}(Y,F) \\ & - (\mu(Y) + I(Y))\, V_{FF} - \frac{1}{2}\sigma(Y)^2 V_{FYY}(Y,F) \end{aligned} \tag{A.3}$$

where $p(Y,F) = -V_F(Y,F)$, $p_F(Y,F) = -V_{FF}(Y,F)$, $p_Y(Y,F) = -V_{FY}(Y,F)$ by smooth pasting conditions. $\phi(Y,F) = \bar{\phi}(Y,F)$ in the break-even by debt region, and $\phi(Y,F) = \frac{(\tau_b-\tau_c)c}{p_F(Y,F)}$ in the equity issuing region. Compare (A.3) with (1.14), if $\bar{\phi}(Y,F) > \frac{(\tau_b-\tau_c)}{p_F(Y,F)}$ within a neighborhood of the boundary in the break-even by debt region, then $-V_{FYY}(Y,F) < p_{YY}(Y,F)$. Then since $p(Y,F) > -V_F(Y,F)$ in the break-even by debt region, this region must be on the "right" side (with higher $\frac{Y}{F}$) of the equity issuing region. Otherwise if $\bar{\phi}(Y,F) < \frac{(\tau_b-\tau_c)}{p_F(Y,F)}$ within a neighborhood of the boundary in the break-even by debt region, the break-even by debt region must be on the "left" side (with lower $\frac{Y}{F}$) of the equity issuing region. However, since equity issuance is positive at the "left" boundary of the equity issuing region, by continuity of security values, the firm should issue more debt when not issuing equity, that is, $\bar{\phi}(Y,F) > \frac{(\tau_b-\tau_c)}{p_F(Y,F)}$. That leads to a contradiction. Therefore, the break-even by debt region must be on the "right" side of the equity issuing region. There is at most one continuous equity issuing region where $p(Y,F) = -V_F(Y,F)$.

Similarly, if there exists a region of (Y,F) such that the first order condition $p(Y,F) = -\frac{V_F(Y,F)}{1-\tau_e}$ holds (dividend distributing region), at the boundary between this region and the break-even by debt region,

$$\begin{aligned} -rV_F(Y,F) =& (1-\tau_e)\left[(1-\tau_c)c - p_F(Y,F)\phi + m\right] + mV_F(Y,F)) - (\phi - mF)V_{FF}(Y,F) \\ & - \left(\mu(Y) + I(Y)\right)V_{FF} - \frac{1}{2}\sigma(Y)^2 V_{FYY}(Y,F) \end{aligned} \tag{A.4}$$

where $p(Y,F) = -\frac{V_F(Y,F)}{1-\tau_e}$, $p_F(Y,F) = -\frac{V_{FF}(Y,F)}{1-\tau_e}$, $p_Y(Y,F) = -\frac{V_{FY}(Y,F)}{1-\tau_e}$ by smooth pasting conditions. $\phi(Y,F) = \bar{\phi}(Y,F)$ in the break-even by debt region, and $\phi(Y,F) = \frac{(\tau_b-\tau_c)c}{p_F(Y,F)}$ in the dividend distributing region. Compare (A.4) with (1.14), if $\bar{\phi}(Y,F) > \frac{(\tau_b-\tau_c)}{p_F(Y,F)}$ within a neighborhood of the boundary in the break-even by debt region, then $-V_{FYY}(Y,F) < p_{YY}(Y,F)$. Then since $p(Y,F) < -\frac{V_F(Y,F)}{1-\tau_e}$ in the break-even by debt region, this region must be on the "left" side of the dividend distributing region. If $\bar{\phi}(Y,F) < \frac{(\tau_b-\tau_c)}{p_F(Y,F)}$ within a neighborhood of the boundary in the break-even by debt region, then $-V_{FYY}(Y,F) > p_{YY}(Y,F)$. Then since $p(Y,F) < -\frac{V_F(Y,F)}{1-\tau_e}$ in the break-even by debt region, this region must be on the "right" side of the dividend distributing region. By continuity of $\bar{\phi}(Y,F)$

and $\frac{(\tau_b-\tau_c)}{p_F(Y,F)}$, a break-even by debt region cannot be on the "right" of one dividend distributing region while being on the "left" of another dividend distributing region. Therefore, there is at most one continuous dividend distributing region where $p(Y,F) = -\frac{V_F(Y,F)}{1-\tau_e}$.

Then we can summarize the regions of financing policies as in Proposition 4.

Proof of proposition 5

By Proposition 4, when $y = \frac{Y}{F}$ is large enough, $\phi(y) = \bar{\phi}(y) < 0$ if $\tau_c < \tau_b$ and $\phi(y) = 0$ if $\tau_c = \tau_b$. Therefore, the firm's leverage converges to zero when it is low enough.

Let y^0 denote the solution of $y = \frac{(1-\tau_c)c+m(1-p(y))}{1-\tau_c-\kappa i}$. If $y_e < \frac{(1-\tau_c)c+m(1-p(y_e))}{1-\tau_c-\kappa i}$, then for $y \in (y_e, y^0)$, $\phi(y) = \bar{\phi}(y) = -\frac{F}{p(y)}\left[(1-\tau_c-\kappa i)y_e - (1-\tau_c)c - m\right] > mF$. The firm's leverage converges to the leverage at y_e. When $y < y_e$, $\phi(y) = \frac{(\tau_c-\tau_b)cF}{yp'(y)} < mF$. The firm's leverage also converges to the leverage at y_e. Therefore, the leverage ratio at y_e is also a local leverage target.

By Proposition 4 and the monotonicity of $\bar{\phi}(y)$, there cannot be other leverage targets.

A.2. The algorithm for solving the model numerically

Here I describe the algorithm for solving the HJB differential equations for the security values $v(y)$ and $p(y)$.

Step 1. Start with a guess of the bankruptcy threshold $\hat{y}_b$.

Step 2. Make a guess of the boundary $\hat{y}_e$ between the equity issuing region and the break-even by debt region, which is larger than y_b.

Step 3. Starting with $\hat{v}(\hat{y}_b) = \hat{v}'(\hat{y}_b) = \hat{p}(\hat{y}_b) = 0$, generate $\hat{v}(y), \hat{p}(y)$ by the following algorithm.

A fourth-order Runge-Kutta-Nystrm algorithm for $v''(y) = \mathscr{G}(y, p(y), p'(y), v(y), v'(y))$ and $p''(y) = \mathscr{H}(y, p(y), p'(y), v(y), v'(y))$:

(1) Let

$$l_1^v = \mathscr{G}(y, p(y), p'(y), v(y), v'(y)) \tag{A.5}$$

$$l_1^p = \mathscr{H}(y, p(y), p'(y), v(y), v'(y)) \tag{A.6}$$

$$v_1' = v'(y) + l_1^v h/2 \tag{A.7}$$

$$p_1' = p'(y) + l_1^p h/2 \tag{A.8}$$

$$v_1 = v(y) + (v'(y) + v_1')/2 \times h/2 \tag{A.9}$$

$$p_1 = p(y) + (p'(y) + p_1')/2 \times h/2 \tag{A.10}$$

where h is a small step size.

(2) Let

$$l_2^v = \mathscr{G}(y + h/2, p_1, p_1', v_1, v_1') \tag{A.11}$$

$$l_2^p = \mathscr{H}(y + h/2, p_1, p_1', v_1, v_1') \tag{A.12}$$

$$v_2' = v'(y) + l_2^v h/2 \tag{A.13}$$

$$p_2' = p'(y) + l_2^p h/2 \tag{A.14}$$

$$v_2 = v(y) + (v'(y) + v_2')/2 \times h/2 \tag{A.15}$$

$$p_2 = p(y) + (p'(y) + p_2')/2 \times h/2 \tag{A.16}$$

(3) Let

$$l_3^v = \mathscr{G}(y + h/2, p_2, p_2', v_2, v_2') \tag{A.17}$$

$$l_3^p = \mathscr{H}(y + h/2, p_2, p_2', v_2, v_2') \tag{A.18}$$

$$v_3' = v'(y) + l_3^v h \tag{A.19}$$

$$p_3' = p'(y) + l_3^p h \tag{A.20}$$

$$v_3 = v(y) + (v'(y) + v_3')/2 \times h \tag{A.21}$$

$$p_3 = p(y) + (p'(y) + p_3')/2 \times h \tag{A.22}$$

(4) Let

$$l_4^v = \mathscr{G}(y + h, p_3, p_3', v_3, v_3') \tag{A.23}$$

$$l_4^p = \mathscr{H}(y + h, p_3, p_3', v_3, v_3') \tag{A.24}$$

$$v'(y + h) = v'(y) + h/6 \times (l_1^v + 2l_2^v + 2l_3^v + l_4^v) \tag{A.25}$$

$$p'(y + h) = p'(y) + h/6 \times (l_1^p + 2l_2^p + 2l_3^p + l_4^p) \tag{A.26}$$

$$v(y + h) = v(y) + h/6 \times (v_1' + 2v_2' + 2v_3' + v'(y + h)) \tag{A.27}$$

$$p(y + h) = p(y) + h/6 \times (p_1' + 2p_2' + 2p_3' + p'(y + h)) \tag{A.28}$$

Here $\mathscr{G}(y, p(y), p'(y), v(y), v'(y))$ and $\mathscr{H}(y, p(y), p'(y), v(y), v'(y))$ are determined by reorganizing the HJB equations in each region.

Then iterate for $y + h$, until y reachs a large enough threshold such that the security values are close enough to their limits for y converging to infinity.

Step 4. Check if $\hat{p}(y)$ converges to $\frac{(1-\tau_b)c+m}{r+m}$. If not, adjust $\hat{y}_e$ and repeat steps 3-4 until convergence.

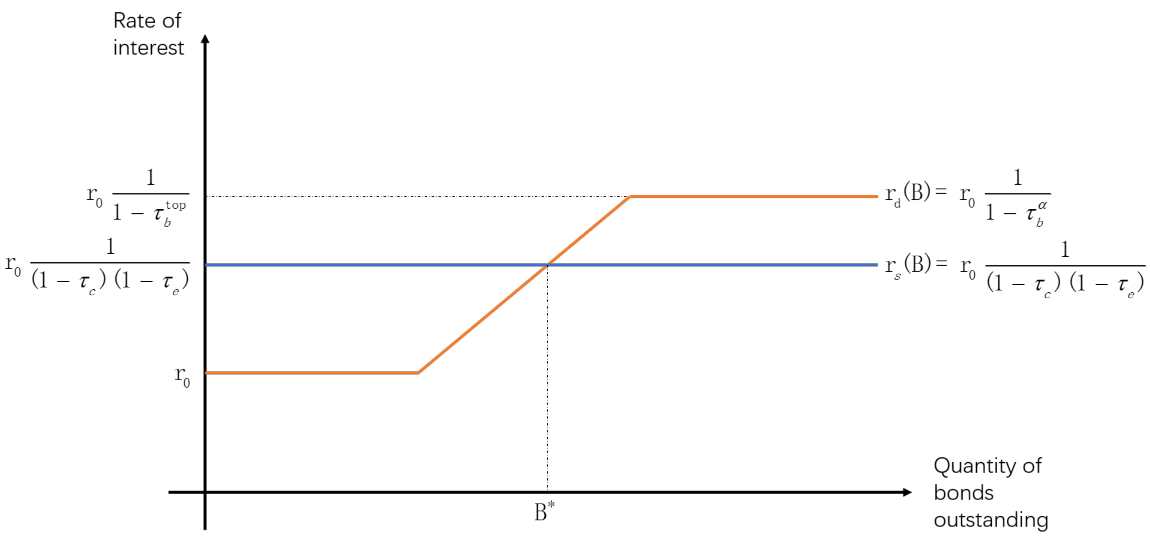
Step 5. Check if $\hat{v}(y)$ converges to $\frac{(1-\tau_e)(1-\tau_c-\kappa i)y}{r-i}$. If not, adjust $\hat{y}_b$ and repeat steps 2-4 until convergence.

Step 6. Check if the results satisfy the equilibrium conditions.

A.3. Extended discussion on Miller (1977)

Miller (1977) describes a market equilibrium where $(1-\tau_c)(1-\tau_e) = 1-\tau_b^{\text{marginal bondholder}}$. In this equilibrium, firms gain no tax benefits on their values. There is no optimal leverage ratio for individual firms but only an equilibrium leverage ratio for the whole corporate sector. Cross-sectional leverage differences are determined by the clientele of firms' bonds with different personal tax rates. The figure below plots all firms' and investors' supply and demand of bonds in this equilibrium following Figure 1 in Miller (1977). There are no frictions except taxes. r_0 is the interest rate of tax-exempt bonds. The upward-sloping part of the demand curve represents that interest rates have to increase to attract investors in higher tax brackets as the amount of debt outstanding grows. Investors with low personal tax rates gain all the surplus.

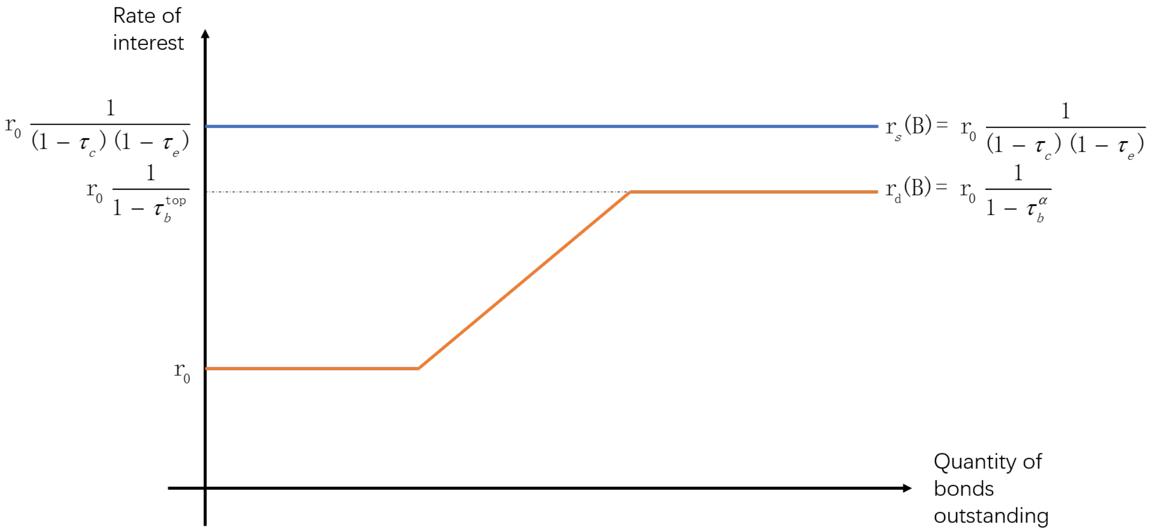
Figure: Market equilibrium in Miller's (1977) framework



Such equilibrium requires shareholders to make tax rates on capital gains small enough by tax concessions. "In the limiting case ... that $(1-\tau_c)(1-\tau_{ps})$ implied a value for τ_{pb}^{α} greater than the top bracket of the income tax, then no interior market equilibrium would be possible." However, empirical measures of effective tax rates on equity income are typically not small enough for the equation to hold without τ_b exceeding the top rates. Then the supply and demand curves become the following .[1]

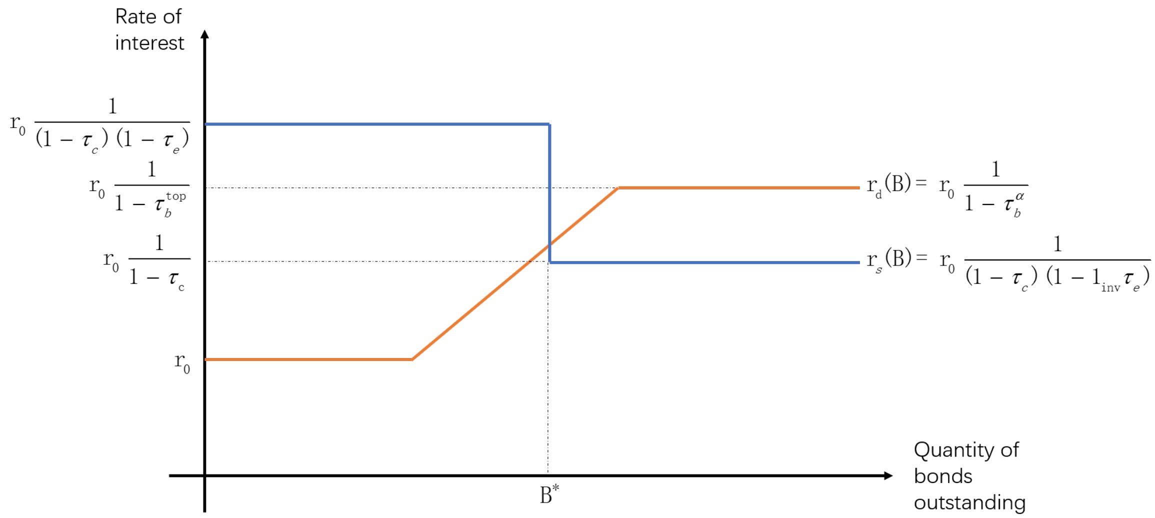
[1]One way to recover an interior equilibrium here is to consider gradient corporate tax rates, which make the supply curve downward sloping.

Figure: No interior equilibrium in Miller's (1977) framework



The different tax consequences of financing investments by debt and leveraged recapitalizations imply a different shape of the supply curve. Therefore, firms supply bonds at rate $r_0\frac{1}{(1-\tau_c)(1-\tau_e)}$ only for financing investments. When recapitalizing for tax shields, firms offer rate $r_0\frac{1}{(1-\tau_c)}$. The figure below plots a revised equilibrium in this manner. In such an equilibrium, the marginal bondholders' personal tax rate can be anywhere between τ_c and $1-(1-\tau_c)(1-\tau_e)$. Firms gain a surplus from debt. Cross-sectional distributions of leverage depend on firms' external financing need.

Figure: Market equilibrium in revised framework



A.4. A model without leverage adjustments (following Leland 1994)

Here I model the key mechanism of this paper into a stylized model without dynamic leverage adjustments following Leland (1994b). A firm earns an exogenous cash flow following a lognormal process, issues debt at time 0, and rolls over the debt. In addition, I assume the firm needs external financing at the beginning and considers personal taxes. I solve the model in closed form and show that the firm's capital structure choice largely depends on the amount of external financing needed at time 0 due to the tax benefit differences between external financing and recapitalizing. When the firm needs no external financing, as in Leland (1994b), opposite to the traditional result without personal taxes, the firm issues no debt if the personal income tax rate on interest payments is no less than the corporate tax rate.

A.4.1. Model setup

Investors and the firm are risk neutral. There exists a risk-free asset paying a constant rate of return r after tax. A firm's before-tax cash flow follows

$$\frac{dY_t}{Y_t} = \mu dt + \sigma dZ_t \tag{A.29}$$

where $\mu < r$. At time 0, the firm needs to finance an investment $I \geq 0$ by issuing debt or equity to earn the cash flow.[2] Assume that $I < \frac{(1-\tau_c)(1-\tau_e)Y}{r-\mu}$, the investment does not exceed the firm's unleveraged value. The firm issues homogenous debt with a coupon rate c and total principal F that matures exponentially at rate $m \geq 0$. It rolls over matured debt until bankruptcy.

There are three taxes at constant rates: a corporate tax at rate τ_c, a personal income tax on bonds at rate τ_b, and a personal income tax on equity at rate τ_e. By a constant rate of personal tax on equity, I am assuming that the firm's distribution strategy and shareholders' tax deferral streategy are fixed over time, so that each dollar available to shareholders are taxed equally.[3] For simplicity, assume the firm holds no cash and distributes all free cash flow as dividends.[4] The firm maximizes the total value of after-tax dividends for shareholders and claims bankruptcy when it is optimal. At bankruptcy, a fraction $\alpha \in [0, 1)$ of the firm's unleveraged after-tax value $v_{unlev}(Y_b) = \frac{(1-\tau_c)(1-\tau_e)Y_b}{r-\mu}$ can be recovered and paid to debtholders, where Y_b denotes pre-tax earnings at bankruptcy.

[2] If $I = 0$, the firm starts with no need for external financing as in Leland (1994b).

[3] Deferring the realizations of personal taxes on equity by stock repurchases or cash holdings can be represented by a lower value of the parameter τ_e, as long as the firm's distribution strategy is fixed over time.

[4] Cite papers here.

A.4.2. Optimal debt issuance

Denote $v(Y)$ as the firm's equity value after dividends or equity issuance for $t \geq 0$ and $v_0(Y)$ as the equity value before dividends or equity issuance at time 0. Let $p(Y)$ be the price of debt with a unit face value that rolls over until bankruptcy, equaling the after-tax value of payments earned by debtholders, and $\tilde{p}(Y)$ be the value of the firm's payments to this debt before personal income tax on bonds. Then $v_0(Y)$ can be written as the sum of time 0 after-tax dividends (with negative value representing equity issuance) and the equity value after dividends or equity issuance $v(Y)$, where $v(Y)$ equals the unleveraged cash flow value $v_{unlev}(Y)$ plus tax benefits for saving corporate tax $\mathcal{TB}_c(Y)$ and personal income tax on equity $\mathcal{TB}_e(Y)$ on the cash flow minus bankruptcy costs $\mathcal{BC}(Y)$ and the value of payments to debtholders $\tilde{p}(Y)F$. At time 0, the firm chooses a face value of debt F to maximize

$$v_0(Y) = \left(1 - \mathbf{1}_{\{p(Y)F-I\geq 0\}}\tau_e\right)[p(Y)F - I] + v(Y)$$

(A.30)

$$= \left(1 - \mathbf{1}_{\{p(Y)F-I\geq 0\}}\tau_e\right)[p(Y)F - I] + v_{unlev}(Y) + \mathcal{TB}_c(Y) + \mathcal{TB}_e(Y) - \mathcal{BC}(Y) - \tilde{p}(Y)F$$

Here $\mathbf{1}_{\{p(Y)F-I\geq 0\}}$ equals 1 if the firm distributes dividends at time 0 and equals 0 if the firm issues equity. Let $\mathcal{TC}_b(Y) = [\tilde{p}(Y) - p(Y)]F$ be the personal income tax costs on bonds, then we can rewrite (A.30) as

(A.31)

$$v_0(Y) = -I - \mathbf{1}_{\{p(Y)F-I\geq 0\}}\tau_e[p(Y)F - I] + v_{unlev}(Y) + \mathcal{TB}_c(Y) + \mathcal{TB}_e(Y) - \mathcal{TC}_b(Y) - \mathcal{BC}(Y)$$

Besides tax benefits and costs on the cash flow, personal income taxes also reduce shareholders' payoff at time 0 by $\tau_e[p(Y)F - I]$ if there is a dividend payment. Therefore, the net tax benefits of debt are reduced if the firm issues more debt than needed for financing the investment.

Solving the value function. Next, I solve each component of $v_0(Y)$ by their HJB equations. The price of debt follows

$$\underbrace{rp(Y)}_{\text{required return}} = \underbrace{(1-\tau_b)c}_{\text{after-tax coupon}} + \underbrace{m[1-p(Y)]}_{\text{rollover gain}} + \underbrace{\mu Y p'(Y) + \frac{1}{2}\sigma^2 Y^2 p''(Y)}_{\text{cash flow evolution}} \tag{A.32}$$

with boundary conditions at infinity $p(\infty) = \frac{(1-\tau_b)c+m}{r+m}$, and at bankruptcy $p(Y_b) = \frac{1}{F}\alpha v_{unlev}(Y_b)$. Then the after-tax value of a par bond is [5]

$$p(Y) = \frac{c(1-\tau_b)+m}{r+m}\left[1-\left(\frac{Y}{Y_b}\right)^{\gamma_1}\right] + \frac{1}{F}\alpha v_{unlev}(Y_b)\left(\frac{Y}{Y_b}\right)^{\gamma_1} \tag{A.33}$$

where

$$\gamma = \frac{-\left(\mu - \frac{1}{2}\sigma^2\right) \pm \sqrt{\left(\mu - \frac{1}{2}\sigma^2\right)^2 + 2\sigma^2(m+r)}}{\sigma^2} \tag{A.34}$$

[5] Here, I assume only coupons are taxed on a bond for tractability, as in Leland (1994b). The value of a unit principal bond that is taxed only on its coupon is $\frac{m+c(1-\tau_b)}{m+yield(1-\tau_b)}$, while that of a bond taxed on its yield is $\frac{m+c-\tau_b yield}{m+yield(1-\tau_b)}$. There is a difference $\frac{(yield-c)\tau_b}{m+yield(1-\tau_b)}$ that makes the simplifying assumption increase the debt price and decrease the tax shield for lower coupon rates. It slightly increases the rollover gain when the firm is close to bankruptcy and the yield is high if $\tau_b - \tau_c > 0$.

The value of payments to a unit face value of debt before personal income taxes $\tilde{p}(Y)$ follows

$$\underbrace{r\tilde{p}(Y)}_{\text{required return}} = \underbrace{c}_{\text{pre-tax coupon}} + \underbrace{m[1-p(Y)]}_{\text{rollover gain}} + \underbrace{\mu Y\tilde{p}'(Y) + \frac{1}{2}\sigma^2 Y^2 \tilde{p}''(Y)}_{\text{cash flow evolution}} \tag{A.35}$$

Substract (A.32) from (A.35) and multiply by F, we get

$$r\mathcal{TC}_b(Y) = \tau_b cF + \mu Y \mathcal{TC}_b'(Y) + \frac{1}{2}\sigma^2 Y^2 \mathcal{TC}_b''(Y) \tag{A.36}$$

with boundary conditions at infinity $\mathcal{TC}_b(\infty) = \frac{\tau_b cF}{r}$, and at bankruptcy $\mathcal{TC}_b(Y_b) = 0$. Then the personal income tax cost on bonds is

$$\mathcal{TC}_b(Y) = \frac{\tau_b cF}{r}\left[1 - \left(\frac{Y}{Y_b}\right)^{\eta_1}\right] \tag{A.37}$$

where

$$\eta = \frac{-\left(\mu - \frac{1}{2}\sigma^2\right) \pm \sqrt{\left(\mu - \frac{1}{2}\sigma^2\right)^2 + 2\sigma^2 r}}{\sigma^2} \tag{A.38}$$

Since $r > \mu$, $m \geq 0$, $\gamma_1 \leq \eta_1 < 0 < 1 < \eta_2 \leq \gamma_2$.

Similarly, the tax benefit from saving corporate taxes is

$$\mathcal{TB}_c(Y) = \frac{\tau_c cF}{r}\left[1 - \left(\frac{Y}{Y_b}\right)^{\eta_1}\right] \tag{A.39}$$

The value of corporate tax savings (A.39) differs from the value of personal income tax cost on bonds (A.37) only by the tax rates, because they are both based on coupon payments. Each dollar of the firm's cash flow is taxed either by the personal income rate τ_b or the corporate rate τ_c, depending on whether it is used for coupon payments.

The bankruptcy cost follows

$$r\mathcal{BC}(Y) = \mu Y \mathcal{BC}'(Y) + \frac{1}{2}\sigma^2 Y^2 \mathcal{BC}''(Y) \tag{A.40}$$

with boundary conditions at infinity $\mathcal{BC}(\infty) = 0$, and at bankruptcy $\mathcal{BC}(Y_b) = (1 - \alpha)v_{unlev}(Y_b)$. Then

$$\mathcal{BC}(Y) = (1-\alpha)v_{unlev}(Y_b)\left(\frac{Y}{Y_b}\right)^{\eta_1} \tag{A.41}$$

The tax benefit from saving personal tax on equity follows [6]

$$r\mathcal{TB}_e(Y) = \tau_e(1-\tau_c)cF + \tau_e m[1-p(Y)]F + \mu Y \mathcal{TB}_e'(Y) + \frac{1}{2}\sigma^2 Y^2 \mathcal{TB}_e''(Y) \tag{A.42}$$

with boundary conditions at infinity $\mathcal{TB}_e(\infty) = \frac{\tau_e F}{r}\left\{(1-\tau_c)c + m[1-p(\infty)]\right\}$, and at bankruptcy $\mathcal{TB}_e(Y_b) = 0$. Then

$$\mathcal{TB}_e(Y) = \tau_e\left[\tilde{p}(Y)F - \mathcal{TB}_c(Y) - \frac{\alpha}{1-\alpha}\mathcal{BC}(Y)\right] \tag{A.43}$$

The firm saves personal income tax for equity holders by reducing cash available to them, that is, payments to debtholders net of corporate tax and payment at bankruptcy.

[6]Here I abstract from the tax differences between payouts and equity issuance for tractability, assuming that all cash flow between the firm and equity holders faces a flat rate of τ_e. I model this difference in the next section.

Substitute (A.32)(A.39)(A.43)(A.37)(A.41) into (A.31) and reorganize. If $p(Y)F-I \geq 0$, then

$$\begin{aligned} v_0(Y) = & \underbrace{-(1-\tau_e)I}_{\text{investment cost}} + \underbrace{\frac{(1-\tau_c)(1-\tau_e)Y}{r-\mu}}_{\text{unleveraged value}} + \underbrace{(\tau_c-\tau_b)(1-\tau_e)\frac{cF}{r}\left[1-\left(\frac{Y}{Y_b}\right)^{\eta_1}\right]}_{\text{tax benefits on cash flow net of costs at issuance}} \\ & \underbrace{-\left[(1-\alpha)+\tau_e\alpha\right]\frac{(1-\tau_c)(1-\tau_e)Y_b}{r-\mu}\left(\frac{Y}{Y_b}\right)^{\eta_1}}_{\text{bankruptcy cost including prepaid tax}} \end{aligned} \tag{A.44}$$

If $p(Y)F-I \leq 0$, then

$$\begin{aligned} v_0(Y) = & -\underbrace{I}_{\text{investment cost}} + \underbrace{\frac{(1-\tau_c)(1-\tau_e)Y}{r-\mu}}_{\text{unleveraged value}} + \underbrace{\left[(1-\tau_b)-(1-\tau_c)(1-\tau_e)\right]\frac{cF}{r}\left[1-\left(\frac{Y}{Y_b}\right)^{\eta_1}\right]}_{\text{tax benefits on cash flow net of costs at issuance}} \\ & \underbrace{+\tau_e F\left\{\frac{(1-\tau_b)c+m}{r+m}\left[1-\left(\frac{Y}{Y_b}\right)^{\gamma_1}\right]-\frac{(1-\tau_b)c}{r}\left[1-\left(\frac{Y}{Y_b}\right)^{\eta_1}\right]\right\}}_{\text{tax savings from rollover}} \\ & \underbrace{-\left\{(1-\alpha)\left(\frac{Y}{Y_b}\right)^{\eta_1}+\tau_e\alpha\left[\left(\frac{Y}{Y_b}\right)^{\eta_1}-\left(\frac{Y}{Y_b}\right)^{\gamma_1}\right]\right\}\frac{(1-\tau_c)(1-\tau_e)Y_b}{r-\mu}}_{\text{bankruptcy cost including prepaid tax}} \end{aligned} \tag{A.45}$$

The tax benefits for generating each dollar of interest expense is $(\tau_c-\tau_b)(1-\tau_e)$ when it is generated by recapitalizing and is $(1-\tau_b)-(1-\tau_c)(1-\tau_e)$ when it is generated by financing investment by debt. The difference $\tau_e(1-\tau_b)$ is due to personal income tax on equity charged on time 0 dividends. In the first case, when $p(Y)F-I \geq 0$, all terms in (A.44) are scaled by $(1-\tau_e)$ – as in the literature about the trapped equity view of dividend taxation,[7] when equity issuance is bounded at 0 and cannot be further

[7]cite papers here

reduced, all the firm's cash flow is subject to personal income tax on equity. Besides the deadweight loss at bankruptcy, there is an additional tax cost on the recovery value because the recovery value is priced in debt and added to the dividends at time 0. In the second case, when $p(Y)F - I \leq 0$, debt reduces the cash flow available to shareholders by both interest expenses and rollover losses, leading to an additional term of personal income tax savings on equity.

When $\tau_c \leq \tau_b$, (A.44) is no larger than the firm's unleveraged value since the tax benefits are negative, so the firm always wants to issue less debt if $F > \frac{I}{p(Y)}$. Then we have the following result

Proposition 12. ***(no recapitalization)*** *If $\tau_b \geq \tau_c$, optimal debt issuance $F^* \leq \frac{I}{p(Y)}$, the firm never issues more debt than needed for financing investments.*

This is because the marginal tax benefit of debt issuance becomes negative when equity issuance drops to 0 and cannot be further reduced. Proceeds from additional debt have to be distributed to equity holders and taxed by τ_e, so the tax benefit only depends on comparing the corporate tax rate to the personal income tax rate on coupons.

Optimal default. The bankruptcy threshold Y_b in (A.44)(A.45) is chosen endogenously such that the firm claims bankruptcy when equity value and it's dirivative to earnings reaches 0, ie.e, $v(Y_b) = 0$ and $v'(Y_b) = 0$. The equity value after time 0 is

$$
\begin{aligned}
\text{(A.46)} \qquad v(Y) =& v_{unlev}(Y) + \mathcal{TB}_c(Y) + \mathcal{TB}_e(Y) - \mathcal{BC}(Y) - \tilde{p}(Y)F \\
=& \underbrace{\frac{(1-\tau_c)(1-\tau_e)Y}{r-\mu}}_{\text{unleveraged value}} + \underbrace{\frac{(1-\tau_e)(\tau_c-\tau_b)cF}{r}\left[1-\left(\frac{Y}{Y_b}\right)^{\eta_1}\right]}_{\text{tax benefits on cash flow}} \\
&- \underbrace{(1-\tau_e)\frac{c(1-\tau_b)+m}{r+m}F\left[1-\left(\frac{Y}{Y_b}\right)^{\gamma_1}\right]}_{\text{flow payments to debtholders}} \\
&+ \underbrace{(1-\tau_e)\alpha v_{unlev}(Y_b)\left[\left(\frac{Y}{Y_b}\right)^{\eta_1} - \left(\frac{Y}{Y_b}\right)^{\gamma_1}\right] - v_{unlev}(Y_b)\left(\frac{Y}{Y_b}\right)^{\eta_1}}_{\text{bankruptcy cost to equity holders'}}
\end{aligned}
$$

By the smooth pasting condition $v'(Y_b) = 0$, we get

$$
\text{(A.47)} \qquad Y_b = \frac{\frac{(\tau_c-\tau_b)c}{r}\eta_1 - \frac{(1-\tau_b)c+m}{r+m}\gamma_1}{\frac{1-\tau_c}{r-\mu}\left[1-(1-\tau_e)\alpha(\gamma_1-\eta_1)-\eta_1\right]}F
$$

When debt is perpetual, i.e., $m = 0$, there is no rollover of debt and $\gamma_1 = \eta_1$, then $Y_b = \frac{-(r-\mu)\eta_1 c}{r(1-\eta_1)}F$, the same as in Leland (1994a). Tax rates does not affect default decisions. When debt has finite maturity, i.e, $m > 0$, then the bankruptcy threshold Y_b increases with τ_b and decreases with τ_e, since debtholders' personal income tax decreases the rollover gain while equity holders' personal income tax increases the rollover gain.

Optimal debt issuance. We can solve the optimal debt issuance F by substituting (A.47) into ((A.44)) and (A.45), then maximize the time-0 value function over F. For simplicity, I focus on the case when debt is perpetual so that $m = 0$. [8] I denote $\widetilde{F}$ as the optimal debt issuance when the firm issues equity at time 0 and refer to it as the opitmal

[8]When $m > 0$, optimal debt issuance cannot be solved in closed form due to the rollover gains term of equity holders' personal tax. I discuss results with rollover and optimal maturity in the next section with endogenous leverage adjustments and leave the analysis without leverage adjustments in the appendix.

financing leverage. When $p(Y)F - I \leq 0$, debt issuance that maximizes (A.45) is [9]

$$\widetilde{F}^* = \min\left\{\widetilde{F}, \frac{I}{p(Y)}\right\} \tag{A.48}$$

where

$$\widetilde{F} = \frac{r(1-\eta)}{-\eta(r-\mu)c}\left[1 - \eta - \frac{(1-\tau_c)(1-\tau_e)}{(1-\tau_b)-(1-\tau_c)(1-\tau_e)}(1-\alpha)\eta\right]^{\frac{1}{\eta}} Y \tag{A.49}$$

The optimal financing leverage is increasing in the corporate tax rate τ_c and the personal income tax rate on equity τ_e, and decreasing in the personal income tax rate on bond τ_b. It coincides with a Leland model setting the constant tax benefit as Miller's formula $(1-\tau_b)-(1-\tau_c)(1-\tau_e)$ and the unleveraged value of the firm as $\frac{(1-\tau_c)(1-\tau_e)}{r-\mu}Y$. Next, I denote $\widehat{F}$ as the optimal debt issuance when the firm distributes dividends at time 0 and refer to it as the opitmal recapitalizing leverage. When $p(Y)F - I \geq 0$, debt issuance that maximizes (A.44) is

$$\widehat{F}^* = \max\left\{\widehat{F}, \frac{I}{p(Y)}\right\} \tag{A.50}$$

where

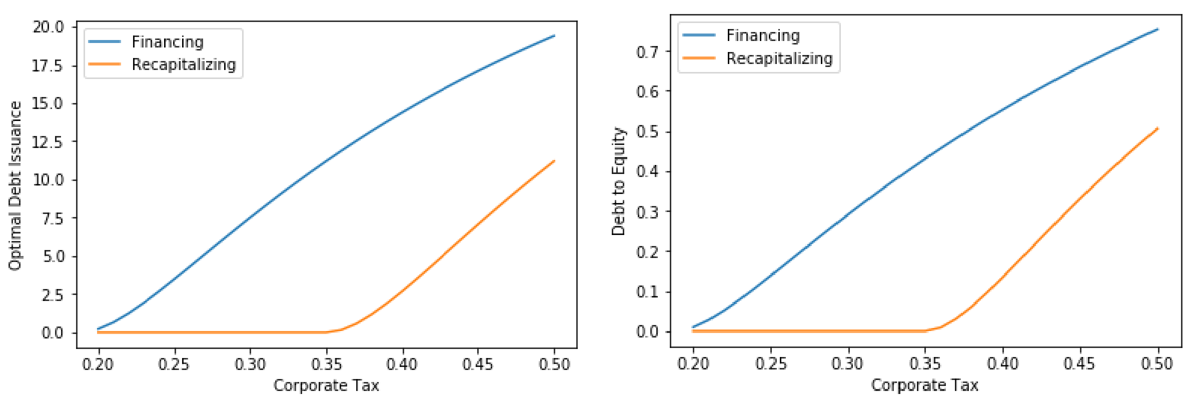
$$\widehat{F} = \frac{r(1-\eta)}{-\eta(r-\mu)c}\left[1 - \eta - \frac{1-\tau_c}{\tau_c-\tau_b}[(1-\alpha)+\tau_e\alpha]\eta\right]^{\frac{1}{\eta}} Y \tag{A.51}$$

The optimal recapitalizing leverage is increasing in the corporate tax rate τ_c and decreasing in the personal income tax rates τ_b and τ_e. Personal income tax on equity becomes a cost rather than benefit for recapitalizing since the recovery value at bankruptcy is priced in

[9] Derivations of optimal debt issuance are in the appendix.

debt and included in the proceeds from debt at time 0, which is paid to equity holders and taxed. The optimal recapitalizing leverage is strictly lower than the optimal financing leverage when $\tau_e > 0$, $\tau_c < 1$, and $(1-\tau_b) > (1-\tau_c)(1-\tau_e)$. The Figure below plots the optimal financing leverage and optimal recapitalizing leverage with different corporate tax rates τ_c and fixed personal tax rates $\tau_b = 35\%$, $\tau_e = 20\%$. When $\tau_c < \tau_b$, $\widehat{F} < 0$ and the firm does not recapitalize. [10] The difference between two leverage targets is largest when the corporate tax rate τ_b is close to the personal tax rate on bond τ_b. Since the corporate tax rates and personal income tax rates are usually close in practice, the model implies that leverage targets for a firm facing an investment problem and a recapitalization problem are very different.

Figure: Leverage targets without leverage adjustments



The firm chooses between the optimal financing leverage $\widetilde{F}^*$ and the optimal recapitalizing leverage $\widehat{F}^*$ to maximize $v_0(Y)$.

Proposition 13. ***(optimal debt issuance)*** *The optimal financing leverage $\widetilde{F}$ is higher than the optimal recapitalizing leverage $\widehat{F}$. Let $\underline{F}$ be the smallest F such that*

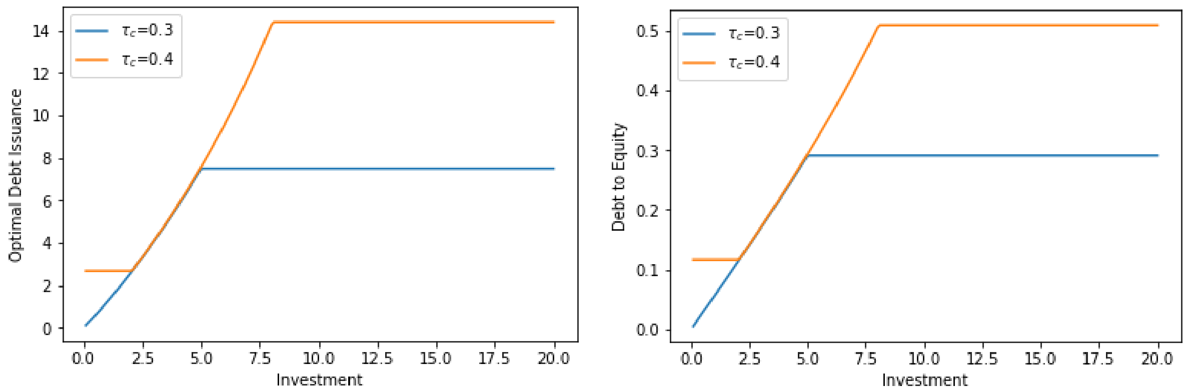
[10]Since the firm starts with no debt, negative debt issuance at time zero is infeasible.

$I = p(Y,F)F$ *if such* F *exists. The firm's optimal debt issuance is*

$$(A.52) \qquad F^* = \begin{cases} \widehat{F} & \text{if } \frac{I}{p(Y,\widehat{F})} \leq \widehat{F} \\ \underline{F} & \text{if } \frac{I}{p(Y,\widehat{F})} > \widehat{F} \text{ and } \frac{I}{p(Y,\widetilde{F})} \leq \widetilde{F} \\ \widetilde{F} & \text{if } \frac{I}{p(Y,\widetilde{F})} > \widetilde{F} \end{cases}$$

The firm chooses the optimal recapitalizing leverage when it is enough to finance the investment. Otherwise, the firm chooses the lowest level of debt that can exactly finance the investment if proceeds from the optimal financing leverage exceeds the financing need for investment, and chooses the optimal financing leverage if it is not enough the finance the investment. The Figure below plots optimal debt issuance with different investment I and fixed personal tax rates $\tau_b = 35\%$, $\tau_e = 20\%$ when the corporate rate is $\tau_c = 30\%$ and $\tau_c = 40\%$. The firm issues debt at the recapitalizing target when required investment I is low, and at the financing target when required investment I is high. Besides the two leverage targets, the firm issues (the minimum level of) debt that exactly meets the investment need without paying dividends or issuing equity when facing a moderate level of investment. Debt issuance is lower than the financing target so issuing debt is cheaper than issuing equity and reducing debt issuance is suboptimal. On the other hand, debt issuance is higher than the recapitalizing target, so issuing more debt and distributing the proceeds as dividends is also suboptimal.

Figure: Optimal Debt issuance without leverage adjustments



A.4.3. Debt policies with existing debt

Debt issuance is bounded at 0 when there is no financing need for investment and the personal tax rate is higher than the corporate tax rate. Now we relax this bound by assuming that the firm has existing debt at time 0 with principal F_0 and the same c and m as the new debt. The initial debt has no covenants and does not restrict the firm's issuance of new debt. Also, we allow I to be negative here, representing the financing need net of internal cash at time 0. When I is negative, there are some internal cash available for payouts or debt repurchase. Let F be the total principal of debt after time 0. Then equity holders' payoff at time 0 becomes $\left(1 - \mathbf{1}_{\{p(Y)(F-F_0)-I\geq 0\}}\tau_e\right)[p(Y)(F-F_0)-I]$ and the time-0 value function is

$$\begin{aligned} v_0(Y) = &- p(Y)F_0 - I - \mathbf{1}_{\{p(Y)(F-F_0)-I\geq 0\}}\tau_e[p(Y)(F-F_0)-I] + v_{unlev}(Y) \\ &+ \mathcal{TB}_c(Y) + \mathcal{TB}_e(Y) - \mathcal{TC}_b(Y) - \mathcal{BC}(Y) \end{aligned} \tag{A.53}$$

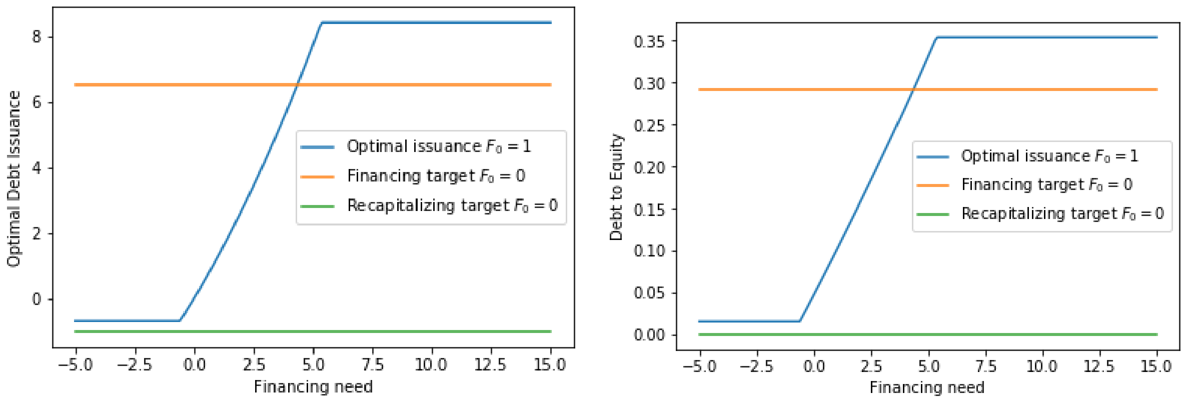
The initial debt decreases the firm's value by its value at the current price. A lower debt price impairs the existing debt holders and benefits the equity holders. Such friction between equity holders and debt holders leads to the debt ratchet effect, making the firm take higher leverage.

Proposition 14. ***(Optimal debt issuance with existing debt)*** *If the firm has existing debt with face value $F_0 > 0$ at the beginning of time 0, the optimal debt issuance $F^{**}(Y, F_0) > F^*(Y) - F_0$. When $\tau_b > \tau_c$ and $I < 0$, the firm repurchases debt if*

$$F_0 < Y \frac{(1-\eta)r}{-\eta(r-\mu)c} \left(\frac{\tau_b - \tau_c}{\tau_b - \tau_c - \frac{\eta}{1-\eta}(1-\tau_c)(1-\tau_e)\alpha} \right)^{-\frac{1}{\eta}} \tag{A.54}$$

When the net financing need $I < 0$, equity is trapped with equity issuance bounded at 0, so all payouts to shareholders are taxed at τ_e. Debt repurchase earns a marginal tax benefit proportional to $\tau_b - \tau_c$. Such benefit dominates the debt ratchet effect if the existing debt F_0 is not too large, leading to a debt repurchase. The figure below plots the optimal debt issuance when $F_0 = 1$, compared to the issuance that adjusts debt from F_0 to the leverage targets $\widetilde{F}$ and $\widehat{F}$ without existing debt. The firm repurchases debt when $I < 0$ and issues debt otherwise. Debt issuance/repurchase are fixed at target levels when I is large/small enough, with both targets higher than targets without exisitng debt. As before, the firm issues debt that exactly meets the financing need for moderate levels of I.

Figure: Optimal Debt issuance with existng debt



APPENDIX B

Appendix for Chapter 2

B.1. Additional tables

Table B.1. Summary statistics of replicated marginal corporate tax rates (1980-2007)

	np_mtr	rw_mtr	np_Blouin	rw_Blouin	rw_Graham
obs	146,579	146,579	157,513	157,513	125,669
mean	27%	26%	29%	28%	29%
std	15%	16%	12%	14%	15%
min	0%	0%	0%	0%	0%
1st quartile	17%	10%	19%	17%	18%
median	34%	34%	33%	34%	34%
3rd quartile	35%	35%	35%	35%	36%
max	51%	51%	51%	51%	51%

Notes. This table provides summary statistics of my replication of marginal corporate tax rate estimates by Blouin, Core, and Guay (2010) and Graham (2000), compared to their original estimated reported in Blouin, Core, and Guay (2010).

Table B.2. Summary statistics of marginal tax benefits for the whole sample

	Ex-ante		Ex-post	
	new	traditional	new	traditional
obs	199,436	199,436	199,436	199,436
mean	-12.9%	-6.32%	-12.24%	-5.66%
std	13.44%	16.41%	14.01%	17.16%
min	-47.5%	-45%	-47.5%	-45%
1st quartile	-21.36%	-16.1%	-22.44%	-18.55%
median	-9.92%	-4.62%	-8%	-2.37%
3rd quartile	-3.36%	2.82%	-2%	5.27%
max	25.26%	25.28%	33.12%	36.72%

Notes. This table reports the summary statistics of my marginal tax benefit measures for the whole sample.

APPENDIX C

Appendix for Chapter 3

C.1. MOC Orders and Off-Exchange Trades

C.1.1. Predominant use of MOC orders in close auctions

Table C.1. NYSE Group Order Type Usage (March 2020): Percentage of Matched Volumes

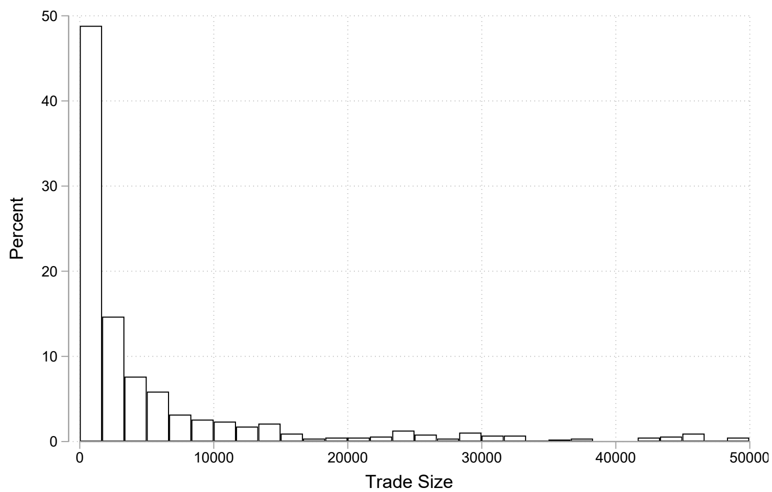
	NYSE	NYSE Arca	NYSE American
Auction	**28.21%**	**5.95%**	**8.07%**
Market-on-Close	16.37%	3.72%	4.61%
Limit-on-Close	3.63%	1.18%	1.59%
Market-on-Open	1.88%	0.55%	0.99%
Limit-on-Open	1.22%	0.49%	0.88%
Closing D-Orders	5.07%	0.00%	0.00%
Closing Offset	0.04%	0.00%	0.00%

Notes: The table reports the percentage of matched total daily volumes constituted by different order types. The table should be interpreted as: in NYSE in Mar 2020, 28.21% of total daily volume happens in auctions, and 16.37% of total daily volume are triggered by market-on-close orders. Source: `https://www.nyse.com/publicdocs/nyse/NYSE_Group_Executed_Order_Type_Usage.xlsx`

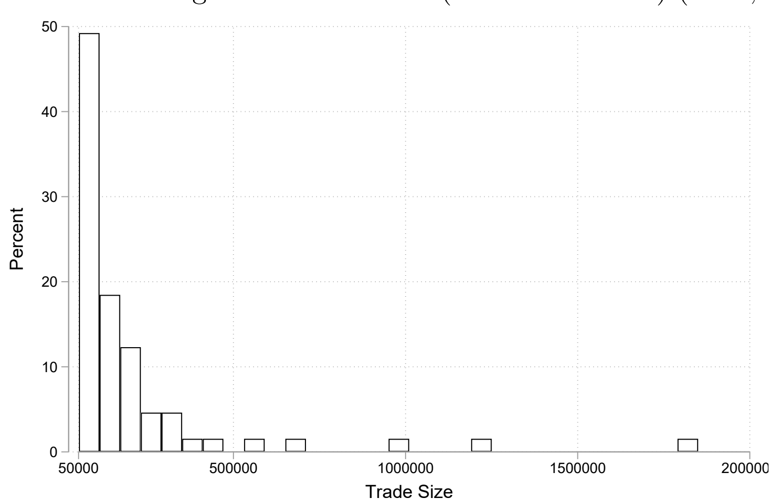
C.1.2. Distribution of off-exchange trade size

Figure C.1. Histogram of off-exchange MOC trade size for stock AAPL in Jan 2018

(a) Histogram of off-exchange MOC trade size (AAPL Jan 2018) ($\leq 50,000$ Shares)



(b) Histogram of off-exchange MOC trade size (AAPL Jan 2018) ($> 50,000$ Shares)



C.2. Proofs

C.2.1. Proof of Proposition 7

The bank chooses $d(y)$ that maximizes its expected trading profits, given the net orders received $y = u_b + \theta\eta v$. The bank solves the following optimization problem

$$\max_d E[v - p|y]d(y) \tag{C.1}$$

when taking the fee φ_b as given. Here α, θ are the fractions of uninformed and informed traders that trade through the bank. Guess that the market maker applies a linear price setting rule $p = \lambda z = \lambda\left[u + \eta v + d(y)\right]$. For simplicity, we write $d(y)$ as d in the following proof.

Knowing that $v \sim N(0, \sigma^2), u \sim N(0, \sigma_u^2)$, the bank's expectation of the asset value is

$$\begin{aligned} E[v|y] &= E[v] + \frac{Cov(v,y)}{Var(y)}(y - E[y]) \\ &= \frac{\theta\eta\sigma^2}{\theta^2\eta^2\sigma^2 + \alpha\sigma_u^2}y \end{aligned} \tag{C.2}$$

And similarly the bank's expectation of net liquidity orders is

$$E[u|y] = \frac{\alpha\sigma_u^2}{\theta^2\eta^2\sigma^2 + \alpha^2\sigma_u^2}y \tag{C.3}$$

Then we can solve for the bank's problem as

$$\max_d E[v-p|y]d = E[v-\lambda\left[u+\eta v+d\right]|y]d$$
$$= \left((1-\lambda\eta)E[v|y]-\lambda E[u|y]-\lambda d\right)d$$
$$= \left(\left[(1-\lambda\eta)\frac{\theta\eta\sigma^2}{\sigma_y^2}-\lambda\frac{\alpha\sigma_u^2}{\sigma_y^2}\right]y-\lambda d\right)d \tag{C.4}$$

where $\sigma_y^2 = \theta^2\eta^2\sigma^2+\alpha\sigma_u^2$. Taking first order conditions, we get

$$d = \frac{1}{2\lambda}\left[(1-\lambda\eta)\frac{\theta\eta\sigma^2}{\sigma_y^2}-\lambda\frac{\alpha\sigma_u^2}{\sigma_y^2}\right]y$$
$$= \left[\frac{1}{2\lambda}\frac{\theta\eta\sigma^2}{\sigma_y^2}-\frac{1}{2}\frac{\theta\eta^2\sigma^2+\alpha\sigma_u^2}{\sigma_y^2}\right]y \tag{C.5}$$

Hence

$$d = K(\lambda)y \tag{C.6}$$

where

$$K(\lambda) = \frac{1}{2\lambda}\frac{\theta\eta\sigma^2}{\sigma_y^2}-\frac{1}{2}\frac{\theta\eta^2\sigma^2+\alpha\sigma_u^2}{\sigma_y^2} \tag{C.7}$$

Next we derive the price setting rule of the market maker and show that it takes the linear form as we guessed. The market maker receives net order

$$z = u+\eta v+d(y) = u_m+(1+K(\lambda))u_b+(1+K(\lambda)\theta)\eta v \tag{C.8}$$

and knows that $v \sim N(0, \sigma^2), u \sim N(0, \sigma_u^2)$. Since the market maker is competitive, the close price equals the market maker's expected value of the asset. Then the close price is

$$p = E[v|z] = \frac{(1+K(\lambda)\theta)\eta\sigma^2}{\left((1+K(\lambda))^2\alpha + (1-\alpha)\right)\sigma_u^2 + (1+K(\lambda)\theta)^2\eta^2\sigma^2} z \tag{C.9}$$

which can be written as

$$p = \lambda z \tag{C.10}$$

where

$$\lambda = \frac{(1+K(\lambda)\theta)\eta\sigma^2}{\left((1+K(\lambda))^2\alpha + (1-\alpha)\right)\sigma_u^2 + (1+K(\lambda)\theta)^2\eta^2\sigma^2} \tag{C.11}$$

Therefore, the equilibrium is characterised by the bank's optimal strategy (C.7) and the market maker's price setting rule (C.11). Substituting λ as a function of K, i.e., Equation (C.11) into Equation (C.7), we can solve for K as a result of a quadratic equation:

$$0 = \sigma_y^2 K^2 + \left(\theta\eta^2\sigma^2 + (2-\theta)\frac{\alpha}{\theta}\sigma_u^2\right) K + \left(\frac{\alpha}{\theta} - 1\right)\sigma_u^2 \tag{C.12}$$

where $\sigma_y^2 = \theta^2\eta^2\sigma^2 + \alpha\sigma_u^2$. Then

$$K = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \tag{C.13}$$

where $A = \sigma_y^2, B = \theta\eta^2\sigma^2 + (2-\theta)\frac{\alpha}{\theta}\sigma_u^2, C = \left(\frac{\alpha}{\theta} - 1\right)\sigma_u^2$.[1] And

$$\lambda = \frac{(1+K\theta)\eta\sigma^2}{\left((1+K)^2\alpha + (1-\alpha)\right)\sigma_u^2 + (1+K\theta)^2\eta^2\sigma^2} \tag{C.14}$$

[1]The other solution of the quadratic equation $\frac{-B-\sqrt{B^2-4AC}}{2A}$ does not optimize the bank's profit here.

Both K, λ are constants that only depend on parameters. The proposition is hence proved. □

C.2.2. Proof of Proposition 8

By substituting previous results into the first term, the expected trading profit can be written as $\lambda(\varphi_b)K^2(\varphi_b)\sigma_y(\varphi_b)^2 \geq 0$, which is non-negative.

When the above sufficient conditions are satisfied, the bank makes a positive profit by setting a fee slightly lower than φ_m and match some liquidity orders. Note that the bank's total profit is negative when φ_b goes to $-\infty$. Since the bank's total profit is a continuous function of φ_b, there exists $\varphi_b < \varphi_m$ such that the break-even conditions holds.

C.2.3. Proof of Proposition 9

We first show that

$$\frac{(1+K\theta)^2}{(1+K)^2\alpha + 1 - \alpha} \geq 1 \tag{C.15}$$

where

$$K = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \tag{C.16}$$

where $A = \sigma_y^2, B = \theta a^2\sigma^2 + (2-\theta)\frac{\alpha}{\theta}\sigma_u^2, C = \left(\frac{\alpha}{\theta} - 1\right)\sigma_u^2$.

If $\theta = \alpha$, then $K = 0$ and the LHS equals 1.

If $\theta > \alpha$, then $K > 0$, the problem is equivalent to

$$\theta^2 K + 2\theta > \alpha K + 2\alpha \tag{C.17}$$

The inequality immediately holds when $\theta^2 \geq \alpha$. When $\theta^2 < \alpha$, the inequality can be written as

$$K < \frac{2(\theta - \alpha)}{\alpha - \theta^2} \tag{C.18}$$

By substituting K as a function of the parameters and rearranging, we can write the inequality as

$$0 < D_1 a^2 \sigma^2 + D_2 \sigma_u^2 \tag{C.19}$$

where

$$D_1 = 2\frac{(\theta - \alpha)\theta}{(\alpha - \theta^2)^2}(\theta^2 - 2\alpha\theta + \alpha) \tag{C.20}$$

$$D_2 = 4\frac{(\theta - \alpha)^2}{(\alpha - \theta^2)^2}\alpha + \frac{1}{\alpha - \theta^2}\left(3\alpha - 2\alpha\theta + \theta^2\right)\left(1 - \frac{\alpha}{\theta}\right) \tag{C.21}$$

Since $D_1 > 0$ and $D_2 > 0$, the inequality holds.

If $\theta < \alpha$, then $K < 0$, the problem can be written as

$$K > \frac{2(\theta - \alpha)}{\alpha - \theta^2} \tag{C.22}$$

By substituting K as a function of the parameters and rearranging, we can write the inequality as

$$0 > D_1 \eta^2 \sigma^2 + D_2 \sigma_u^2 \tag{C.23}$$

where

$$D_1 = 2\frac{(\theta-\alpha)\theta}{(\alpha-\theta^2)^2}(\theta^2 - 2\alpha\theta + \alpha) \tag{C.24}$$

$$D_2 = \frac{\theta-\alpha}{(\alpha-\theta^2)^2}\left[3\frac{\alpha^2}{\theta} - 6\alpha^2 + 2\alpha\theta + 2\alpha\theta^2 - \theta^3\right] \tag{C.25}$$

Since $D_1 < 0$ and $D_2 < 0$, the inequality holds.

Then

$$\xi = \frac{(1+K\theta)^2\eta^2\sigma^2}{\left[(1+K)^2\alpha + 1 - \alpha\right]\sigma_u^2} \geq \frac{\eta^2\sigma^2}{\sigma_u^2} = \xi_{\text{no bank}} \tag{C.26}$$

Hence $MSE_{\text{with bank}} \leq MSE_{\text{no bank}}$. Having the bank conducting the service improves price informativeness.

C.2.4. Proof of Proposition 10

Since the MSE measure is strictly decreasing in ξ, it is sufficient to show that $\frac{\partial\xi(\delta,\zeta)}{\partial\zeta} \geq 0$, where $\delta = \frac{\theta}{\alpha}$ and $\zeta = \frac{\alpha\sigma_u^2 + \theta\eta^2\sigma^2}{\sigma_u^2 + \eta^2\sigma^2}$.

First, notice that $\zeta = \alpha\frac{\sigma_u^2 + \delta\eta^2\sigma^2}{\sigma_u^2 + \eta^2\sigma^2}$, so if we rewrite $\xi(\delta,\zeta)$ as $\xi(\delta,\alpha)$, $\frac{\partial\xi(\delta,\zeta)}{\partial\zeta} = \frac{\sigma_u^2 + \delta\eta^2\sigma^2}{\sigma_u^2 + \eta^2\sigma^2}\frac{\partial\xi(\delta,\alpha)}{\partial\alpha}$. Therefore, showing $\frac{\partial\xi(\delta,\zeta)}{\partial\zeta} \geq 0$ is equivalent to showing that $\frac{\partial\xi(\delta,\alpha)}{\partial\alpha} \geq 0$. We write $K(\alpha,\delta)$

as K for simplicity.

(C.27)

$$\frac{\partial\xi(\alpha,\delta)}{\partial\alpha} = \frac{\eta^2\sigma^2}{\sigma_u^2}\frac{\partial}{\partial\alpha}\frac{[1+\alpha\delta K(\alpha,\delta)]^2}{[1+K(\alpha,\delta)]^2\alpha+1-\alpha}$$

$$= \frac{\eta^2\sigma^2}{\sigma_u^2}\frac{1}{[(1+K)^2\alpha+1-\alpha]^2}\left\{2(1+\alpha\delta K)\left(\delta K+\alpha\delta\frac{\partial K}{\partial\alpha}\right)\left[(1+K)^2\alpha+1-\alpha\right]\right.$$

(C.28) $$\left. -(1+\alpha\delta K)^2\left[2K+K^2+2\alpha(1+K)\frac{\partial K}{\partial\alpha}\right]\right\}$$

Since $\frac{\eta^2\sigma^2}{\sigma_u^2}\frac{(1+\theta K)}{[(1+K)^2\alpha+1-\alpha]^2} > 0$, $\frac{\partial\xi(\alpha,\delta)}{\partial\alpha}$ takes the same sign as

(C.29)

$$\tilde{\xi} = 2\left(\frac{\theta}{\alpha}K+\theta\frac{\partial K}{\partial\alpha}\right)\left[(1+K)^2\alpha+1-\alpha\right]-(1+\theta K)\left[2K+K^2+2\alpha(1+K)\frac{\partial K}{\partial\alpha}\right]$$

(C.30)

$$= \theta K^3+(2\theta-1)K^2+2\left(\frac{\theta}{\alpha}-1\right)K+2\left[(\theta-1)\alpha K+\theta-\alpha\right]\frac{\partial K}{\partial\alpha}$$

Rewrite (C.13) as a function of α and δ and take its partial derivative to α, we get

(C.31) $$\frac{\partial K}{\partial\alpha} = -K\frac{(2\alpha\delta^2\eta^2\sigma^2+\sigma_u^2)K+\delta\eta^2\sigma^2-\sigma_u^2}{\sqrt{\left[\alpha\delta\eta^2\sigma^2+(\frac{2}{\delta}-\alpha)\sigma_u^2\right]^2-4(\alpha^2\delta^2\eta^2\sigma^2+\alpha\sigma_u^2)(\frac{1}{\delta}-1)\sigma_u^2}}$$

(C.32) $$= -K\frac{(2\frac{\theta^2}{\alpha}\eta^2\sigma^2+\sigma_u^2)K+\frac{\theta}{\alpha}\eta^2\sigma^2-\sigma_u^2}{\sqrt{\left[\theta\eta^2\sigma^2+(\frac{2}{\theta}-1)\alpha\sigma_u^2\right]^2-4(\theta^2\eta^2\sigma^2+\alpha\sigma_u^2)(\frac{\alpha}{\theta}-1)\sigma_u^2}}$$

(C.33) $$= -\frac{K}{\alpha}+\sigma_u^2K\frac{K+\frac{2}{\theta}}{\sqrt{\left[\theta\eta^2\sigma^2+(\frac{2}{\theta}-1)\alpha\sigma_u^2\right]^2-4(\theta^2\eta^2\sigma^2+\alpha\sigma_u^2)(\frac{\alpha}{\theta}-1)\sigma_u^2}}$$

Then (C.30) can be written as

$$\tilde{\xi} = \theta K^3 + (2\theta - 1)K^2 + 2\left(\frac{\theta}{\alpha} - 1\right)K - 2\left[(\theta-1)\alpha K + \theta - \alpha\right]\frac{K}{\alpha}$$
$$+ 2\left[(\theta-1)\alpha K + \theta - \alpha\right]\sigma_u^2 K \frac{K + \frac{2}{\theta}}{\sqrt{\left[\theta\eta^2\sigma^2 + (\frac{2}{\theta} - 1)\alpha\sigma_u^2\right]^2 - 4(\theta^2\eta^2\sigma^2 + \alpha\sigma_u^2)(\frac{\alpha}{\theta} - 1)\sigma_u^2}}$$

(C.34)

$$= (1 + \theta K)K^2$$

(C.35)

$$+ 2\left[(\theta-1)\alpha K + \theta - \alpha\right]\sigma_u^2 K \frac{K + \frac{2}{\theta}}{\sqrt{\left[\theta\eta^2\sigma^2 + (\frac{2}{\theta} - 1)\alpha\sigma_u^2\right]^2 - 4(\theta^2\eta^2\sigma^2 + \alpha\sigma_u^2)(\frac{\alpha}{\theta} - 1)\sigma_u^2}}$$

Since $K \geq -\frac{1}{2}$ and $\theta \in [0,1]$, the first term $(1+\theta K)K^2$ is non-negative and is positive when $K \neq 0$. The second term has the same sign as $\left[(\theta-1)\alpha K + \theta - \alpha\right]K$. Next we show that it is positive when $K \neq 0$.

When $\theta = 1$, it becomes $(\theta - \alpha)K > 0$. When $\theta < 1$, let $\tilde{K} = \frac{\theta - \alpha}{(1-\theta)\alpha}$. Since K and $\tilde{K}$ have the same sign as $\theta - \alpha$, $\left[(\theta-1)\alpha K + \theta - \alpha\right]K$ is positive as long as $|\tilde{K}| > |K|$. Since K is the larger root of (C.12), $\sigma_y^2 > 0$, $\theta\eta^2\sigma^2 + (\frac{2}{\theta} - 1)\alpha\sigma_u^2 > 0$ and $\tilde{K}$ has the same sign as K, $|\tilde{K}| > |K|$ as long as

$$0 < K\left[\sigma_y^2\tilde{K}^2 + \left(\theta\eta^2\sigma^2 + (2-\theta)\frac{\alpha}{\theta}\sigma_u^2\right)\tilde{K} + \left(\frac{\alpha}{\theta} - 1\right)\sigma_u^2\right] \tag{C.36}$$

$$= K\left[\frac{\left(\frac{\theta}{\alpha} - 1\right)^2}{(1-\theta)^2}(\theta^2\eta^2\sigma^2 + \alpha\sigma_u^2) + \left(\theta\eta^2\sigma^2 + (2-\theta)\frac{\alpha}{\theta}\sigma_u^2\right)\frac{\frac{\theta}{\alpha} - 1}{1-\theta} + \left(\frac{\alpha}{\theta} - 1\right)\sigma_u^2\right] \tag{C.37}$$

$$= \frac{K(\theta - \alpha)}{(1-\theta)^2}\left[\left(\frac{\theta^2}{\alpha} - 2\theta + 1\right)\frac{\theta}{\alpha}\eta^2\sigma^2 + \left(\frac{\theta}{\alpha} - 2 + \frac{1}{\theta}\right)\sigma_u^2\right] \tag{C.38}$$

The inequality holds since $K(\theta-\alpha) > 0$, $\left(\frac{\theta^2}{\alpha} - 2\theta + 1\right) > 0$, and $\left(\frac{\theta}{\alpha} - 2 + \frac{1}{\theta}\right) > 0$ when $\alpha, \theta \in (0,1)$. Then (C.33) is non-negative and is positive when $K \neq 0$. Therefore, $\frac{\partial \xi(\alpha,\delta)}{\partial \alpha} \geq 0$ and the inequality is strict when $\theta \neq \alpha$.

C.2.5. Proof of Proposition 11

The expected trading profit before fees of informed traders is

$$\Pi_x = \mathbb{E}[\mathbb{E}[(v-p)\eta v|v]] \tag{C.39}$$

$$= \eta\sigma^2 \frac{\left((1+K)^2\alpha + (1-\alpha)\right)\sigma_u^2}{\left[(1+K)^2\alpha + (1-\alpha)\right]\sigma_u^2 + (1+K\theta)^2\eta^2\sigma^2} \tag{C.40}$$

The expected trading profit before fees of liquidity traders who trade directly with the exchange is

$$\Pi_{um} = \mathbb{E}[\mathbb{E}[(v-p)u_m|u_m]] \tag{C.41}$$

$$= -(1-\alpha)\sigma_u^2 \frac{(1+K\theta)\eta\sigma^2}{\left[(1+K)^2\alpha + (1-\alpha)\right]\sigma_u^2 + (1+K\theta)^2\eta^2\sigma^2} \tag{C.42}$$

The expected trading profit before fees of liquidity traders who trade with the bank is

$$\Pi_{ub} = \mathbb{E}[\mathbb{E}[(v-p)u_b|u_b]] \tag{C.43}$$

$$= -\alpha\sigma_u^2 \frac{(1+K)(1+K\theta)\eta\sigma^2}{\left[(1+K)^2\alpha + (1-\alpha)\right]\sigma_u^2 + (1+K\theta)^2\eta^2\sigma^2} \tag{C.44}$$

Note again that the equilibrium results with no bank coincides with the results when we impose $K = 0$, that is, the bank do not trade. Then we compare the profits of the traders with and without the bank and get the following results.

Informed traders' expected trading profit before fees in the "no bank" equilibrium coincides with the profit when $K = 0$, that is

$$\Pi_{x,\text{no bank}} = \eta\sigma^2 \frac{\sigma_u^2}{\sigma_u^2 + \eta^2\sigma^2} \tag{C.45}$$

Following Proposition 9, we have $\Pi_x < \Pi_{x,\text{no bank}}$.

Similarly, liquidity traders' expected trading profit before fees in the "no bank" equilibrium is

$$\Pi_{u,\text{no bank}} = -\sigma_u^2 \frac{\eta\sigma^2}{\sigma_u^2 + \eta^2\sigma^2} \tag{C.46}$$

For traders who still trade with the exchange in the equilibrium with the bank, they are better of if and only if

$$\Pi_{um} > (1-\alpha)\Pi_{u,\text{no bank}} \tag{C.47}$$

That is equivalent to

$$\lambda < \lambda_{\text{no bank}} = \frac{\eta\sigma^2}{\sigma_u^2 + \eta^2\sigma^2} \tag{C.48}$$

When $\theta < \alpha$ and $K < 0$, $1 > 1 + K\theta > (1+K\theta)^2 > (1+K)^2\alpha + (1-\alpha)$, then $\Pi_{um} < (1-\alpha)\Pi_{u,\text{no bank}}$.

When $\theta > \alpha$ and $K > 0$, substitute K from Proposition 7 into the inequality (C.47) and reorganize, we get

$$(\theta - 3\alpha)\sigma_u^2 < 2\theta\eta^2\sigma^2 \tag{C.49}$$

For traders who trade with the bank in the equilibrium with the bank, they are better off if and only if

$$\Pi_{ub} > \alpha \Pi_{u,\text{no bank}} \tag{C.50}$$

When $\theta > \alpha$ and $K > 0$, $(1+K)(1+K\theta) > (1+K\theta)^2 > (1+K)^2\alpha + (1-\alpha)$, then $\Pi_{ub} < \alpha \Pi_{u,\text{no bank}}$.

When $\theta < \alpha$ and $K < 0$, substitute K from Proposition 7 into the inequality (C.50) and reorganize, we get

$$(1-\theta)(1+K\theta)\eta^2\sigma^2 + [1+\theta-2\alpha+(\theta-\alpha)K]\sigma_u^2 > 0 \tag{C.51}$$