Shareholders' Expected Recovery Rate and Underleverage Puzzle

Daniel Kim*

May 27, 2018

Abstract

I address underleverage puzzle by relaxing Absolute Priority Rule. Shareholders' strategic default action, whose severity is determined by shareholders' expected recovery rate, acts as a "negative" commitment device. Thus, firms' optimal leverage decreases over shareholders' expected recovery rate. This channel helps to match empirically observed leverage and default probability. Structural estimation yields 19.8% of expected bankruptcy cost and 7% of shareholders' expected recovery rate, both of which are in line with the previous literature' finding. Time-series subsample analysis reveals that shareholders' expected recovery rate increased and bankruptcy cost decreased after shareholder-friendly Bankruptcy Reform Act was passed in 1978. Furthermore, consistent with the empirical literature, my subsample and firm-level estimation results show that firm size is a good positive proxy for shareholders' expected recovery rate and can potentially explain why underleverage puzzle seems to be pronounced among large firms.

*The Wharton School, University of Pennsylvania. I am deeply indebted to my dissertation committee: Joao Gomes, Christian Opp, Nikolai Roussanov, Luke Taylor and Amir Yaron for their insightful comments, guidance and support. I would like to thank Lorenzo Garlappi, Wei Wang, David Skeel, Lin Shen, Jinyuan Zhang, participants at Trans-Atlantic Doctoral Conference 2018 and Wharton PhD Finance seminar for their valuable comments. I would also like to thank the Rodney L. White Center for Financial Research and Jacobs Levy Equity Management Center for financial support on this project

1 Introduction

The Trade-off theory is, arguably, the most important theory in corporate finance. However, it was empirically rejected, dubbed as underleverage puzzle, because the empirically observed bankruptcy cost is too low to explain empirically observed corporate leverage. Most of exiting studies assumed that Absolute Priority Rule (APR) holds. In this paper, I allow APR to be violated by letting shareholders recover non-negative amount upon bankruptcy and address underleverage puzzle.

Sequence of historical events in the U.S. made the nature of its bankruptcy system shareholder-friendly and thus made it easy for APR to be violated. Prior to the nineteenth century, consistent with the common understanding and bankruptcy laws in other countries, APR always hold. However, in late nineteenth century, series of bankruptcies in railroad industry forced bankruptcy court to change its view on APR violation: the courts were concerned about a possible meltdown of public transit if bankruptcies were handled according to APR. For the sake of public interest, a court managed to involve various parties, including managers and shareholders, and opened the door for possible APR violations. Over the course of following years, this shareholder-friendly practice slowly spread to other industries whose bankruptcies do not necessarily deteriorate public interest. Continuing this trend, in 1978, Bankruptcy Reform Act was passed to further strengthen APR violations.

Accordingly, APR violations in the U.S. are more common than typically believed. Wickes, private company in retail industry, filed for bankruptcy on April 24th, 1982 and emerged from the bankruptcy on Sept 21st, 1984. Pure size of the company made the case very complicated: it was the largest non-railroad company to date to emerge from bankruptcy and it involved 150,000 creditors with total outstanding debt amount in \$1.6 billion. Sanford Sigoloff (chairman and CEO) was able to pull off a corporative environment among shareholders, managers, creditors and employees and the company successfully emerged from the bankruptcy in much shorter time than many believed. According to the Wall Street Journal (September 24th, 1984), all parties agreed to the violation of APR: common shareholders were given \$57M (4% of the total distribution) even though creditors were not fully paid (\$246M less than what they were owed). The Washington Post and the New York Times hailed the case as "textbook" treatment of the original intent of the bankruptcy law. However, Wickes case could be due to idiosyncratic factors and could effectively make the Wickes' outcome externally invalid and thus my paper attempts to fill this gap.

Key economic question is, how does APR violation help researchers to address underleverage puzzle and eventually validate the Trade-off theory? In this paper, I focus on one type of APR violations: shareholders recover non-negative amount even though creditors were not paid in full. When shareholders expect to recover higher amount upon bankruptcy, shareholders optimally choose to strategically default sooner than later and that implies higher default probability. Anticipating shareholders' strategic default action, debt becomes more costly and thus firms optimally choose to lower leverage. In other words, shareholders' recovery rate upon bankruptcy acts as a "negative" commitment device and ex-ante optimal leverage decreases as a result. This channel allows to match empirically observed leverage with reasonable bankruptcy cost. Furthermore, this helps to estimate shareholders' expected recovery rate and quantitatively answer how likely *average* firms *expect* APR violation to occur.

In order to illustrate the above point, I form a structural model and estimate bankruptcy cost (α) and shareholders' expected recovery rate (η). Full sample analysis yields that α is 19.8% and η is 7.0%. These results are interesting for the following two reasons. First, surprisingly, shareholders expect to recover 7% of firm value upon bankruptcy as opposed to 0% as typically assumed in the standard capital structure model. This clearly illustrates that firms expect APR to be violated. Second, α of 19.8% is closely in line with extant literature's estimates.

I have five contributions at large. First, a number of existing studies relating leverage to bankruptcy cost assume that APR holds (or equivalently $\eta = 0$). However, I show that data imply that firms do not expect APR to hold. Moreover, I show that relaxing APR helps to partially address underleverage puzzle. Furthermore, consistent with a number of empirical literature, I show that shareholder-friendly bankruptcy act, BRA 1978, increased η.

Second, this is the first paper to structurally estimate η that is implied in prices and accounting data of non-bankruptcy firms. Traditional papers estimated the ex post recovery rate of shareholders based on a small sample of bankrupt firms. While these traditional papers are instructive, such results can potentially suffer from various bias such as sample selection bias and small sample bias. I perform my analysis by directly estimating ex ante expected recovery rate of shareholders that are implied in observable prices and accounting data by examining a broad cross-section of non-bankrupt firms. Interestingly, I show that such bias in η might not be too large.

Third, I speak to another dimension of underleverage puzzle that has not received much attention yet. Both Graham (2000) and Lemmon and Zender (2001) found that underleverage tends to be more pronounced among large firms that are typically deemed to face low bankruptcy cost. Via both subsample and firm-level estimations, I show that η increases over firm size and thus could potentially explain why underleverage is more pronounced among large firms.

Fourth, although growing literature has found η to be important, because η is unobservable, they have to rely on observable proxies. Due to lack of guidance on proxies' validity, the literature uses wide range of different proxies. Through subsample analysis and firm-level analysis, this paper attempts to fill this gap. Consistent with the literature practice, I show that firm size is a good positive proxy for η.

Fifth, I augment dynamic capital structure model by allowing shareholders to recover $\eta \in [0, 1]$ fraction of remaining firm value. More specifically, upon bankruptcy, firms incur bankruptcy cost $\alpha \in [0, 1]$, shareholders recover η and creditors recover the remainder $1 - \eta - \alpha$. This modification is realistic because I focus on publicly listed firms. These firms almost always attempt to renegotiate upon bankruptcy[1] and thus their shareholders expect to recover non-zero value if firms go bankrupt. Current model is different from Fan and Sundaresan (2000)-type renegotiation model that endogenizes η by exogenously setting shareholders' bargaining power. Although there is monotonic relation between η and shareholders' bargaining power, there are three major differences that make the current model more suitable for structural estimation than Fan's. Fan used bankruptcy cost α as a bargaining surplus between creditors and shareholders. Thus, Fan's model implies that 1) η is a fixed fraction of α and 2) firms do not incur any bankruptcy cost in equilibrium. My model does not impose restriction 1) and allows data to speak to it. 2) is hardly true as empirical literature (e.g. Andrade and Kaplan (1998)) estimated that firms, which end up renegotiating upon bankruptcy, still incur non-zero bankruptcy cost. Accordingly, the current model allows firms to incur bankruptcy cost even when shareholders and creditors renegotiate. Lastly, η is easier to find an empirical counterpart than more abstract term such as shareholders' bargaining power and thus makes it easier to validate estimation results.

[1]According to LoPucki bankruptcy database, 97.5% of firms in their sample file for Chapter 11.

For careful quantitative exercise, I conduct structural estimation. Based on marginal-tax rates that John Graham provides, I estimate more up-to-date tax rates and show how it can partially address underleverage puzzle. Moreover, as default probability is the key part of the story and identification strategy, I attempt to match default probability. Based on the past literature (Hackbarth et al. (2015), Garlappi et al. (2008) and Garlappi and Yan (2011))'s finding that equity price is sensitive to η, I attempt to match CAPM-β for more accurate η estimation. Lastly, I run different types of structural estimations and compare results. I first assume that firms are homogeneous and attempt to structurally estimate the representative firm's characteristics. Then, in order to address issues that could arise due to heterogeneity in firms, I use two approaches. First, I divide the sample based on typically-used proxies for η and run subsample analysis. Second, similar to Glover (2016), I run firm-level estimation and report its potential limitation.

The rest of the paper is structured as follows. Section 2 discusses in detail the sequence of events in the U.S. that allowed APR to be violated. Section 3 develops the model. Section 4 discusses the main hypothesis and identifying moments. Section 5 explains data construction process. For full-sample and subsample estimation, Section 6 discusses estimation procedure and presents results. For firm-level estimation, Section 7 discusses estimation procedure and presents related results. Lastly, Section 8 concludes.

Literature Review The first strand of literature is on *underleverage puzzle*. According to trade-off theory, a firm optimally chooses a leverage at a point where marginal cost (bankruptcy cost) and marginal benefits (interest tax shield) are balanced. Using various approaches, the literature (e.g. Altman (1984), Andrade and Kaplan (1998), Davydenko et al. (2012), van Binsbergen et al. (2010)) estimated the bankruptcy cost to be between 6.9% and 20%. However, researchers (e.g. Miller (1977), Graham (2000)) found that empirically-observed bankruptcy cost is too low to justify empirically observed leverage. In response to this concern, Almedia and Philippon (2007) used counter-cyclicality of financial distress to address the puzzle. Alternatively, by allowing firms to experience modest financial distress cost prior to the actual bankruptcy, Elkamhi et al. (2012) addressed it. By allowing creditors to recover fraction of *levered* firm value as opposed to *unlevered* firm value (which was their way to model reorganization), Ju et al. (2005) addressed it. Bhamra et al. (2010) (intertemporal macroeconomic risk) and Chen (2010) have attempted to use macro economic risk to address the same puzzle. More recently, Glover (2016) estimated the expected bankruptcy cost to be much larger (45%) by matching leverage and attributed a sample selection bias as a possible reason behind such a low empirical estimate.

By forcing firms to roll-over fixed fraction of debt as opposed to letting them optimally refinance, Reindl et al. (2017) shows that bankruptcy cost is reflected in the market value of newly rollovered debt and therefore in the net distribution to equityholders. By matching equity price and estimating default threshold based on put option pricing data, Reindl et. al. estimated bankruptcy cost to be 20%. Although Reindl et al.'s estimate is similar to mine, we differ in a few major areas. I allow APR to be relaxed, firms in my paper issue perpetuity debt (thus no need to roll over) until it finds itself optimal to upward restructure and shareholders determine the optimal time of bankruptcy.

Second, there is growing literature, both empirical and theoretical, on shareholders' expected recovery rate upon bankruptcy. In violation of APR, shareholders recover non-negative value upon bankruptcy because shareholders can threaten to exercise a few options [2]. Credibility of these threats is the best illustrated in Eastern Airline's bankruptcy case (year 1989), which is arguably the most notorious case for shareholders to exercise these options at the expense of creditors. As Weiss and Wruck (1998) showed, Eastern Airline's shareholders fully exercised their options and destroyed the firm value by 50% during the 69 months-long bankruptcy process. Being aware of chance of shareholders' hostile actions and lengthy and costly bankruptcy process, it is reasonable for creditors to accept shareholders' renegotiating terms, especially when firms are financially distressed. This naturally allows shareholders to recoup non-zero residual value upon default.

In support of the above claim, several empirical papers (Franks and Torous (1989), Betker (1995), Eberhart et al. (1990), Weiss (1990) and Bharath et al. (2007)) found that average shareholders recover non-zero value upon bankruptcy. However, I believe that their measures could be biased in two ways. First, bias could arise because firms with small η tend to default more often than those with large η. The second source of bias is due to how it was measured. The extant literature typically estimates shareholders' recovery rate by using security prices that most closely postdate the firms' emergence from bankruptcy. However, not every firm successfully emerges from bankruptcy. Thus, studying η only among firms that have successfully emerged from bankruptcy could potentially bias η's estimate. My structural estimation is immune from these critiques.

[2] 1) an option to take risky actions (asset substitution), 2) an option to enter costly chapter 11, 3) an option to delay chapter 11 process if entered and 4) an option not to preserve tax loss carryfowards (for asset sales).

Shareholders' non-zero recovery rate, thus violation of APR, has become more common in the US in part thanks to Bankruptcy Reform Act 1978 (see LoPucki and Whitford (1990)). Noting an importance of shareholders' non-zero recovery rate, strategic debt service model was first modeled by Fan and Sundaresan (2000) and then adopted in a number of recent papers (Davydenko and Strebulaev (2007), Garlappi et al. (2008), Garlappi and Yan (2011), Hackbarth et al. (2015), Boualam et al. (2017)). Hackbarth et al. (2015) recently studied the act's impact on equity price. However, there is insufficient study on how much shareholders *expect* to recover upon bankruptcy especially when its bankruptcy is highly unlikely and I fill this gap.

Third, Hackbarth et al. (2015) used drop in CAPM-β as an indirect evidence to support that Bankruptcy Reform Act 1978 increased η. However, this evidence holds true only when everything else are kept constant. This calls for a structural model in order to determine what has caused a drop in CAPM-β. My results imply that η did increase after the law was passed even after accounting for other changes in firm characteristics and confirm Hackbarth et. al's.

Fourth, the current paper is related to vast literature on the relation between tax and leverage. Graham (1999) used panel data to document that cross-sectional variation in tax status affected debt usage. As summarized in Graham (2003), it is important to consider non-debt tax shield, in addition to debt-related tax shield, in calculating firms' MTR and Graham (1996a), Graham (1996b) and Graham (1998) show how to estimate those for each firm at given point in time. Moreover, as noted in Miller (1977), in studying the Trade-off theory, it is important to incorporate personal income tax and dividend tax. In the current paper, I follow the literature to estimate the tax rates for each firm at given time.

Fifth, the literature empirically found that shareholders' non-zero recovery rate has minimal impact on credit spreads across countries (Davydenko and Franks (2008)) nor in the U.S. (Davydenko and Strebulaev (2007)). When leverage choice is exogenous, the model typically implies that high η should lead to higher credit spread due to shareholders' strategic action, which is disadvantageous against creditors. However, when firms internalize higher cost of debt, firms optimally choose smaller leverage. Thus, endogenous leverage choice could dampen η's impact. As a result, readers should not interpret empirically-observed muted response on credit spreads as η being small or not important.

Lastly, Green (2018) studies how valuable restrictive debt covenants is in reducing the agency costs of debt. As the author's focus was on restructuring, he modeled firms' default decision as random event. On the contrary, I took firms' strategic default decision more seriously and study how it impacts firms' financing. Although I do not explicitly model covenant in my model, looser covenant can be matched to higher η and could have the same effect on firms' ex-ante behavior such as leverage and default probability.

2 Bankruptcy Law in the U.S.

In this section[3], I discuss sequence of historical events in the U.S. that eventually led to more frequent violation of APR relative to other countries.

Prior to the nineteenth century, the bankruptcy system in the U.S. was very similar to the counterpart in the U.K. and it was administrative in nature. Bankrupt firms were almost always liquidated, its shareholders did not recover any value and managers were let go. Consequently, APR always hold and shareholders were never a part of the bankruptcy process.

However, there has been a dramatic turn of events due to series of bankruptcies in railroad industry in late nineteenth century. This event prompted a court to step in and rescue them for the sake of public interest in an effective transportation system. The court formed equity receivership to run the firm in bankruptcy. Equity receivership comprised of the managers of the insolvent firm and the investment banks that had served as underwriters when the firm sold stock and debt securities to the public. Investment banks helped to set up committees that represent the interest of shareholders and bondholders. It was natural for investment banks to be part of the bankruptcy process because, as past securities underwriters, they were already familiar with security holders. By the end of the nineteenth century, J. P. Morgan and a small group of other Wall Street banks figured prominently in most of bankruptcy cases.

However, it seemed that shareholder-friendly nature of bankruptcy in the U.S. had come to an end when Chandler Act 1938 was passed. In an attempt to protect widely scattered bondholders and cater to populist hostility against investment banks ignited by the Great Depression (1929-1939), Security and Exchange Commission (SEC), a champion of APR, helped to devise a Chapter X under Chandler Act. Chapter X called for an independent

[3]Most of contents in this section are based on Skeel (2001)

trustee, required strict compliance with APR, and gave the SEC a pervasive oversight role. Chapter X seemed to be a perfect bankruptcy venue for publicly held firms because an alternative venue, Chapter XI, was seen as unsuitable: publicly held firms had significant amount of secured debt and Chapter XI did not permit debtors to restructure secured debt. However, Chandler Act did not impose any restriction on access to Chapter XI, which was meant to be used for mom-and-pop firms and small corporate debtors, and this seemingly naive oversight opened the door for large corporate debtors. In fact, in Chapter XI, the debtor's managers retained control, APR was not required, and the SEC's role was minimal thus Chapter XI was clearly better choice for corporate debtors. More popular usage of Chapter XI and less usage of Chapter X had two significant implications. First, contrary to SEC's intention, the nature of bankruptcy in the U.S. stayed shareholder-friendly and made APR violations possible. Second, SEC's role, strong proponent of APR, in bankruptcy process was greatly reduced and was ultimately removed under a new bankruptcy law: Bankruptcy Reform Act (BRA) 1978.

Chandler Act was considered complicated and vague (Posner (1997) and King (1979)). For this reason, large creditors and bankruptcy lawyers pushed for a reform in the bankruptcy code and BRA was passed in 1978. However, due to long legislative history of the BRA (more than a decade) and the complexity of the codification, it was hard to foresee all the effects of BRA. Section 6.3.2 discusses how the literature differs in their assessment of BRA's impact and quantitatively validates their claims.

3 Model

Similar to the existing literature, I follow standard EBIT-based capital structure models (see e.g.Goldstein et al. (2001)) and assume that the earnings of a firm are split between a coupon, promised to creditors in perpetuity and a dividend, paid to shareholders after tax. Shareholders of each firm make three types of corporate financing decisions: (1) they have the right to default at the time of their choice; (2) they decide when to refinance the debt; and (3) they decide on the amount of debt to be issued at each refinancing. Shareholders exercise their default option if earnings drop below a certain earnings level, called the default threshold.

Because my innovation centers on what happens at bankruptcy, let us first discuss how the extant literature treat it. Under Leland (1994)-type model, shareholders do not receive any amount upon bankruptcy. Thus, firms optimally choose to continue operating under contractual coupon amount until equity value becomes 0. Then, firms cease to exist and are forced to liquidate the remaining firm value. On the contrary, Fan and Sundaresan (2000) models renegotiation between shareholders and creditors and this implies non-negative recovery amount for shareholders upon bankruptcy. Under this model, firms continue operating with contractual coupon amount until cash flow reaches the endogenously-determined threshold. As soon as cash flows reaches the threshold from above, debt becomes equity-like and creditors receive a fixed fraction of cash flow. This fraction is determined based on Nash Game where both parties' outside options are payouts upon liquidation. However, creditors resume receiving the original contractual coupon amount as soon as cash flow increases back up to the threshold. Thus, under this world, firms never cease to exist in equilibrium. There is no empirical counterpart to such a temporarily convertible bond. Moreover, the model uses bankruptcy cost as bargaining surplus between creditors and shareholders. this implies that bankruptcy cost is never realized in equilibrium and shareholders recovery rate is positively proportional to bankruptcy cost.

This paper proposes an alternative model that does not require temporarily convertible bond. I characterize bankruptcy by bankruptcy cost (α) and shareholders' recovery share (η). More specifically, upon bankruptcy, creditors receive $1 - \eta - \alpha$[4] and shareholders recover η of the remaining unlevered firm value. Contrary to Leland, the model allows shareholders to recover non-zero value. Contrary to Fan and Sundaresan, firms can potentially incur bankruptcy cost even when they enter renegotiation. Lastly, rather than exogenously imposing positive relation between bankruptcy cost and shareholders' recovery rate, I allow data to speak to it.

3.1 Setup

Aggregate cash flow $X_{A,t}$ and firm i's cash flow $X_{i,t}$ follow a GBM as follows:

$$\frac{dX_{A,t}}{X_{A,t}} = \mu_A dt + \sigma_A dW_t^A$$
$$\frac{dX_{i,t}}{X_{i,t}} = (\mu_i + \mu_A)dt + \beta_i \sigma_A dW_t^A + \sigma_i^F dW_{i,t}^F$$

[4] This naturally imposes a restriction that $\eta + \alpha <= 1$.

The pricing kernel is exogenously set as:

$$\frac{d\Lambda_t}{\Lambda_t} = -rdt - \varphi_A dW_t^A$$

Under the risk-neutral measure, the cash flow process evolves according to:

$$\frac{dX_{i,t}}{X_{i,t}} = \hat{\mu}dt + \sigma_{i,X}d\hat{W}_{i,t}$$

where $\hat{W}_{i,t}$ is Brownian motion under risk neutral probability measure, $\hat{\mu}_i = \mu_i + \mu_A - \beta_i\sigma_A\varphi_A$ and $\sigma_{i,X} = \sqrt{(\beta_i\sigma_A)^2 + (\sigma_i^F)^2}$. In order to guarantee the convergence of the expected present value of X_t, I impose the usual regularity condition $r - \hat{\mu}_i > 0$. For notational convenience, I drop i in the rest of the document.

3.2 Solutions

First, τ_c denotes tax on corporate earning, τ_i denotes tax on interest income and τ_d denotes tax on equity distributions. For a simpler exposition, this paper uses the following notations:

$$(1-\tau_{cd}) \equiv (1-\tau_c)(1-\tau_d)$$
$$\tau_{cdi} \equiv (1-\tau_i) - (1-\tau_{cd})$$

For an arbitrary value for X_D, X_U and C, I first derive the debt value. Debt is a contingent claim to an after-tax interest payment. Thus, debt value $D(X)$ satisfies the following ODE:

$$\frac{1}{2}\sigma_X X^2 D'' + \hat{\mu}XD' + (1-\tau_i)C = rD$$

Boundary conditions are

$$D(X_D) = (1-\alpha-\eta)\frac{(1-\tau_{cd})X_D}{r-\hat{\mu}}$$
$$D(X_U) = D(X_0)$$

Closed form solution for debt value is:

$$D(X_t) = \frac{(1-\tau_i)C}{r} + A_1 X_t^{\lambda_+} + A_2 X_t^{\lambda_-}$$

where

$$\lambda_{\pm} = \left(\frac{1}{2} - \frac{\hat{\mu}}{\sigma_X^2}\right) \pm \sqrt{\left(\frac{1}{2} - \frac{\hat{\mu}}{\sigma_X^2}\right)^2 + \frac{2r}{\sigma_X^2}}$$

where A_1 and A_2 are:

$$\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} X_D^{\lambda_+} & X_D^{\lambda_-} \\ X_U^{\lambda_+} - X_0^{\lambda_+} & X_U^{\lambda_-} - X_0^{\lambda_-} \end{bmatrix}^{-1} \begin{bmatrix} (1-\alpha-\eta)\frac{(1-\tau_{cd})X_D}{r-\hat{\mu}} - \frac{(1-\tau_i)C}{r} \\ 0 \end{bmatrix}$$

Similarly, for an arbitrary value for X_D, X_U and C, equity value is:

$$E(X_t) = \sup_{\tau^D} \mathbb{E}^{\mathbb{Q}} \left[\int_0^{\tau^D} e^{-rs}(1-\tau_{cd})(X_t - C)ds + e^{-r\tau^D} \cdot E(X_D) \right]$$

where $\tau^D \equiv \inf\{t : X_t \leq X_D\}$.

Here, it is important to note that the above tries to maximize equity value for given coupon amount C. This implies that "optimal" default decision τ^D is made without internalizing default decision's impact on cost of debt and leverage. For example, if default decision was made after internalizing its decision's impact on cost of debt, true optimal default decision is not to default at all, i.e. $\tau^D = \infty$. In other words, firm never choose to default and this effectively makes expected bankruptcy cost zero. As a result, firms choose to max out their leverage to enjoy tax shield benefit. However, this is possible only when shareholders commit to constantly supplying cash by issuing equity even when firms' earning is significantly low. This is economically unfeasible and unrealistic and thus I make an assumption that "optimal" default decision was made without regard to its impact on cost of debt and leverage.

Again, following a contingent claims approach, we have:

$$\frac{1}{2}\sigma_X X^2 E'' + \hat{\mu} X E' + (1-\tau_{cd})(X - C) = rE$$

Boundary conditions are:

$$E(X_D) = \frac{\eta(1-\tau_{cd})X_D}{r-\hat{\mu}}$$

$$E(X_U) = \frac{X_U}{X_0}[(1-\phi)D(X_0) + E(X_0)] - D(X_0)$$

Analytical solution for $E(X_t)$ is:

$$E(X_t) = \frac{1-\tau_{cd}}{r-\hat{\mu}}X_t - \frac{(1-\tau_{cd})C}{r} + B_1 X_t^{\lambda_+} + B_2 X_t^{\lambda_-}$$

where B_1 represents additional benefit for being allowed to upward restructure and B_2 represents additional benefit for being allowed to default. Thus, $B_1 > 0$ and $B_2 > 0$ where

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} X_D^{\lambda_+} & X_D^{\lambda_-} \\ X_U^{\lambda_+} - \frac{X_U}{X_0} X_0^{\lambda_+} & X_U^{\lambda_-} - \frac{X_U}{X_0} X_0^{\lambda_-} \end{bmatrix}^{-1}$$

$$\begin{bmatrix} \frac{(1-\tau_{cd})C}{r} + (\eta - 1)\frac{(1-\tau_{cd})X_D}{r-\hat{\mu}} \\ \left(\frac{X_U}{X_0}(1-\phi) - 1\right)\left(A_1 X_0^{\lambda_+} + A_2 X_0^{\lambda_-} + \frac{(1-\tau_i)C}{r}\right) + \frac{X_U}{X_0}\left(\frac{(1-\tau_{cd})}{r-\hat{\mu}}X_0 - \frac{(1-\tau_{cd})C}{r}\right) - \left(\frac{(1-\tau_{cd})}{r-\hat{\mu}}X_U - \frac{(1-\tau_{cd})C}{r}\right) \end{bmatrix}$$

The last remaining step is to solve for an optimal coupon C, upward restructuring point X_U and default threshold X_D. C and X_U are determined at time 0 (initial point or refinancing point) by solving the following maximization problem:

$$[C, X_U] = \arg\max_{C^*, X_U^*} \left(E(X_0; C^*, X_U^*) + (1 - \phi_D)D(X_0; C^*, X_U^*)\right)$$

where X_D is determined based on the following smooth pasting conditions (see the heuristic derivation of smooth pasting condition in Appendix A

$$\lim_{X_t \downarrow X_D} E'(X_t) = \frac{\eta(1-\tau_{cd})}{r - \hat{\mu}}$$

A few points are worth noting here. First, X_D can be smaller than C, i.e. firms are allowed to costlessly issue equity. Second, the conditions above guarantee that when shareholders choose the time of default, their objective is to maximize the default option implicit in levered equity value. Third, as emphasized by Bhamra et al. (2010), due to fluctuations in firm cash flows and the assumed cost of restructuring, the firm's actual leverage drifts away from its optimal target. In the model, the firm is at its optimally chosen leverage ratio only at time 0 and subsequent restructuring dates.

Rewriting the above objective function yields:

$$[C, X_U] = \arg\max_{C^*, X_U^*} \left\{ \frac{1-\tau_{cd}}{r-\hat{\mu}} X_0 + \underbrace{\frac{\tau_{cdi} - \phi_D(1-\tau_i)}{r} C^* + ((1-\phi_D)A_{1c} + B_{1c}) C^{*(1-\lambda_+)} \left(\frac{X_0}{X_U^*}\right)^{\lambda_+}}_{\text{Benefit}} \right.$$

$$\left. + \underbrace{((1-\phi_D)A_{2c} + B_{2c}) C^{*(1-\lambda_-)} \left(\frac{X_0}{X_D}\right)^{\lambda_-}}_{\text{Cost}} \right\}$$

where

$$A_{1c} = \frac{A_1 X_U^{\lambda_+}}{C^{1-\lambda_+}}$$
$$A_{2c} = \frac{A_2 X_D^{\lambda_-}}{C^{1-\lambda_-}}$$
$$B_{1c} = \frac{B_1 X_U^{\lambda_+}}{C^{1-\lambda_+}}$$
$$B_{2c} = \frac{B_2 X_U^{\lambda_-}}{C^{1-\lambda_-}}$$

Here, the first term in benefit represents the tax benefit at the current coupon rate C^* and the second benefit represents additional tax benefit multiplied by risk-neutral restructuring probability. Cost shows value loss (bankruptcy cost plus future tax benefit) multiplied by risk-neutral default probability.

In their decision to default, shareholders weigh the benefits of holding on to their equity rights and all future dividends and recovery value against the costs of honoring debt obligations while the firm is in financial distress. As η increases and so the trade-off shifts and leads to earlier exercise of the option to default.

It is worth noting two special cases. Setting $\eta = 0$ yields Leland (1994)-type model where only liquidation is a possible bankruptcy outcome. Setting $\alpha = 0$ yields Fan and Sundaresan (2000)-type model where only reorganization with zero bankruptcy cost is a possible bankruptcy outcome.

3.2.1 Moments of Interest

This section summarizes formula for each term of interest.

First, a term for book leverage is:

$$\frac{D(X_0)}{D(X_0) + E(X_0)}$$

In the above, I assume that book value of equity and debt is value of equity and debt at time 0 when firms choose optimal leverage. I decided to match book value ratios as they are often the focus of financing decisions (see Graham et al. (2015)). This naturally allows to focus on debt ratios at refinancing points and thus shows that I do not address underleverage puzzle in aggregate level as pointed out in Bhamra et al. (2010).

Second, based on Harrison (1985), a default probability under physical measure is:

$$DP(X_t)$$
$$= \begin{cases} \Phi\left(\frac{\log\left(\frac{X_D}{X_t}\right)-(\tilde{\mu}-\sigma_X^2/2)T}{\sigma_X\sqrt{T}}\right)+\left(\frac{X_t}{X_D}\right)^{1-2(\tilde{\mu})/\sigma_X^2}\Phi\left(\frac{\log\left(\frac{X_D}{X_t}\right)+(\tilde{\mu}-\sigma_X^2/2)T}{\sigma_X\sqrt{T}}\right) & \text{if } X_t \geq X_D \\ 1 \quad \text{Otherwise} \end{cases}$$

where $\tilde{\mu} = \mu_A + \mu$. Here, because the empirical counterpart is a default probability over the next one year and I use quarterly time unit in the model, I set T to 4 to make data and model-implied moments compatible.

Third, I discuss formula for CAPM-β. A term for return is:

$$\begin{aligned} dR_t &= \frac{dE(X_t)+(1-\tau_{cd})(X_t-C)dt}{E(X_t)} \\ &= \left(\frac{(1-\tau_{cd})(X_t-C)}{E(X_t)}+\frac{E'(X_t)X_t}{E(X_t)}(\mu+\mu_A)+\frac{1}{2}\frac{E''(X_t)X_t^2}{E(X_t)}\sigma_X^2\right)dt \\ &+\frac{E'(X_t)X_t}{E(X_t)}(\beta\sigma_A dW_t^A+\sigma_F dW_t^F) \end{aligned}$$

Let x_t^A be a log of aggregate earning X_t^A. Then,

$$x_t^A - x_t^0 = \mu_A t + \sigma_A W_t^A$$

Using this, a term for CAPM-β is:

$$\text{CAPM-}\beta = \frac{1}{dt}\mathbb{E}_t[dx_t^A dR_t]/\frac{1}{dt}var_t[dx_t^A] = \frac{E'(X_t)X_t}{E(X_t)}\beta$$

Fourth, PE ratio is defined as:

$$\log\left(\frac{E(X_t)}{X_t}\right)$$

3.3 Leverage and Default Probability

This subsection discusses how book leverage and default probability help to identify my key parameters: α and η. In order to clearly see the intuition, I temporarily disallow upward restructuring and check the closed form solutions. Then, I allow upward restructuring in the actual estimation and numerically show that the same intuition still carries through in Figure 2.

3.3.1 Optimal Coupon and Book Leverage

Because book leverage monotonically increases over C and term for C is more intuitive to study, I study how C varies over α and η in this subsection. An optimization problem to solve for C is as follows:

$$C = \arg\max_{C^*} \left\{ \frac{1-\tau_{cd}}{r-\hat{\mu}} X_0 + \underbrace{\frac{\tau_{cdi} - \phi_D(1-\tau_i)}{r} C^*}_{\text{Benefit}} + \underbrace{\left(\frac{X_0}{X_{DC}}\right)^{\lambda_-} ((1-\phi_D)A_{2c} + B_{2c}) C^{*(1-\lambda_-)}}_{\text{Cost}} \right\}$$

where

$$X_{DC} = \frac{X_D}{C} = \frac{r-\hat{\mu}}{r} \frac{-\lambda_-}{1-\lambda_-} \frac{1}{1-\eta} \tag{1}$$

The closed form solution for optimal coupon C is

$$C = \underbrace{\left[\frac{\tau_{cdi} - \phi_D(1-\tau_i)}{r}\right]^{-1/\lambda_-}}_{\text{Tax Shield Benefit}^{-1/\lambda_-}} \cdot \underbrace{\frac{X_0}{X_{DC}}}_{\text{1/Default Threshold}} \cdot \underbrace{[-(1-\lambda_-)((1-\phi_D)A_{2c} + B_{2c})]^{1/\lambda_-}}_{\text{Loss}^{1/\lambda_-}} \tag{2}$$

where $(1-\phi_D)A_{2c}+B_{2c}$ is the loss of firm value upon bankruptcy, normalized by coupon C. As a reminder, note that $\lambda_- < 0$.

The first term represents the tax shield benefit adjusted for debt issuance cost. Intuitively, higher tax shield implies higher C. The denominator of the second term shows that C decreases as shareholders strategically determine high threshold X_{DC}. High X_{DC} implies high default probability thus high expected default cost and low optimal C. The third term represents the loss of firm value upon bankruptcy adjusted for debt issuance cost. High loss of firm value implies low C.

Now, let us discuss how C relates to α and η. The term above can be approximately written in terms of α and η when ϕ_D is set to 0. The intuition below is valid even when ϕ_D is set to some positive value.

$$C \propto \underbrace{(1-\eta)}_{\text{1/Default Threshold}} \underbrace{\left(\frac{\alpha}{1-\eta}\right)^{1/\lambda_-}}_{\text{Loss}^{1/\lambda_-}} \tag{3}$$

The above expression immediately shows that high α implies high value loss thus lower optimal C. High η implies high X_{DC}, which in turn implies high default probability for fixed C. Simultaneously, high η implies high X_{DC}, which in turn implies high value loss upon bankruptcy for fixed C. Taken together, C decreases over η and thus book leverage decreases over η. Lastly, power term, $1/\lambda_-$ determines how sensitive coupon is to loss. Coupon is much more sensitive to loss when default probability is more likely (low expected earning growth or high volatility).

3.3.2 Default Threshold and Default Probability

According to the default probability formula shown in Section 3.2.1, for given parameters other than η and α, there is monotonic relation between default probability and X_D (default threshold). Thus, studying how default probability varies over η and α is almost equivalent to studying how X_D varies over η and α. Interesting relation arises because $X_D = X_{DC}C$ where X_{DC} and C can potentially vary differently over η and α. Now, let us look at a term for X_D:

$$\begin{aligned}\frac{X_D}{X_0} &= \left[\frac{\tau_{cdi} - \phi_D(1-\tau_i)}{r}\right]^{-1/\lambda_-} \cdot [-(1-\lambda_-)((1-\phi_D)A_{1C} + \phi_D A_{2C})]^{1/\lambda_-} \\ &\propto \underbrace{\left(\frac{\alpha}{1-\eta}\right)^{1/\lambda_-}}_{\text{Loss}^{1/\lambda_-}} \end{aligned} \tag{4}$$

where I set ϕ_D to 0 in the last $\propto$. For given C, high η implies high X_{DC} as shown in Equation (1). As explained in the previous subsection, increase in η increases both default probability and value loss. Thus, C has to decrease sufficiently enough to offset high expected default cost driven by increase in both default probability and value loss. Thus, decrease in C more than offsets the increase in X_{DC}. As a result, X_D decreases over η and so does default probability. In other words, conditioned on default probability, leverage decreases over η and this illustrate my key economic channel.

As shown in Equation (1), α does not impact X_{DC}. But high α is associated with high loss of firm value upon bankruptcy thus decreases C. Taken together, as α increases, X_D decreases and thus implies lower default probability. One interesting point to note here is when $\alpha = 0$, expected bankruptcy cost is zero thus default probability stays constant over η

3.4 Bankruptcy Cost and Shareholders' Recovery Rate

In the model, firms do not incur bankruptcy costs prior to declaring bankruptcy. In reality, firms typically incur bankruptcy costs prior to the event of bankruptcy due to variety of factors such as reputation costs, asset fire sales, loss of customer or supplier relationships, legal and accounting fees, and costs of changing management. Moreover, the costs of bankruptcy outside of default are borne directly by equity holders, whereas bankruptcy costs are not directly borne by shareholders in the model. Even though shareholders do not directly incur bankruptcy cost in the model, shareholders indirectly experience costs: as bankruptcy cost increases, debt becomes more costly and shareholders internalize higher debt cost.

Similarly, in the model, shareholders recover only upon default. In reality, prior to declaring bankruptcy, some shareholders can potentially enjoy the benefit of control right by, for example, opportunistically restructuring to change covenants (see Green (2018))[5]. To the extent that shareholders' opportunistic behavior make debt more costly and shareholders internalize higher debt cost, the model captures ex-ante changes in shareholders' behavior. Thus, shareholders' recovery rate η in the model captures such benefits in addition to explicit ex-post recovery value.

On the related note, as Reindl et al. (2017) mentioned, presence of debt covenants could make it infeasible to assume that firms only default when it is ex-post optimal for shareholders. My model and estimation results are adequate as long as debt covenants do not bind or firms optimally choose a debt with covenants that are effectively ex-post optimal for shareholders. In the latter case, η again captures the nature of deb covenants.

4 Hypothesis Development and Identification

Main contribution of this paper is to study how relaxing $\eta = 0$ restriction changes firms' optimal debt choice. To that end, this paper forms a null hypothesis as follows:

H_0: η is 0

[5] For example, fallen angel firms delay refinancing relative to always-junk firms because loose covenants allow shareholders to transfer wealth from creditors.

In the first subsection, I discuss in detail how leverage and default probability help to identify η and α. In the next subsection, I list additional moments that help to identify other parameters.

4.1 α and η

In this subsection, we discuss how book leverage and default probability help to identify (α, η) for given μ, σ^F and β . As discussed in the previous subsection, default probability decreases over η. In order to offset decrease in default probability, α has to decrease to match a given default probability. Thus, infinite number of η and α that matches a given default probability should be downward sloping on η-α space as illustrated in Figure 1 where η-α locus (dotted-curve) matches default probability at 4.02%. Similarly, leverage decreases over η and α thus locus (solid line) that matches leverage of 0.2758 is downward sloping on η-α space.

Restricting η to zero and matching only leverage implies $\alpha = 0.24$, an intercept on α-axis. If we allow η to be non-zero, we can better match both leverage and default probability. Moreover, it helps to imply α that is more in-line with empirical counterpart, which is between 6.9% to 20%.

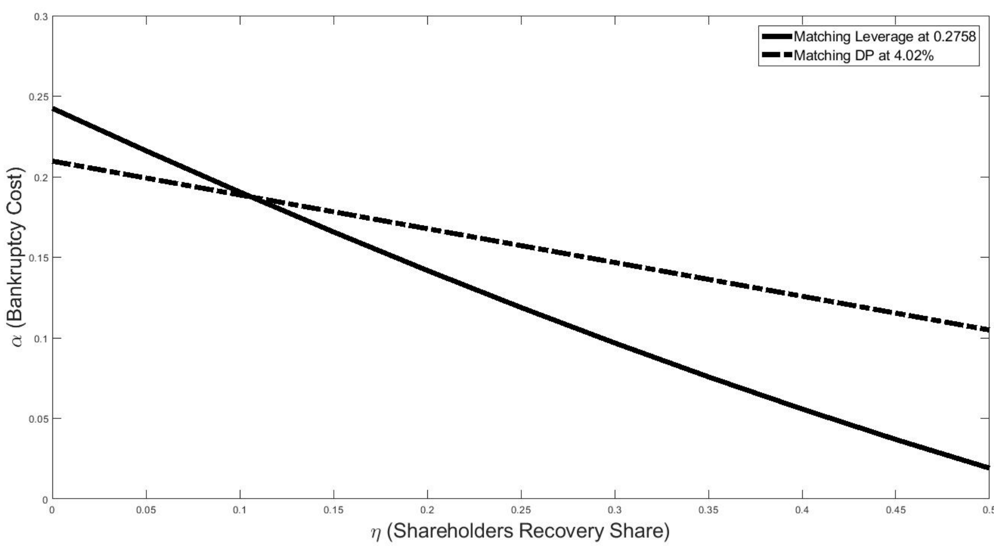

Figure 1: η vs α region using aggregate mean of firm-level parameter estimates.

It is important to note that leverage-locus and default probability-locus have different slopes. This can be easily seen by comparing Equation (2) and (4). The difference between these two terms is 1/Default Prob and this term would differentiate the slope of leverage-locus and default probability-locus. Thus, default probability provides additional information beyond what leverage provides in identifying η and α. As long as default probability plays a role in determining optimal leverage, this is a very general result. There could be a case where two curves do not intersect in the identification region due to other parameter estimates (μ, β, σ^F) that determine curves' horizontal and vertical intercepts. In such cases, default probability and leverage will not be properly matched and implies that the model is rejected by the data.

4.2 Moment Selection

This subsection now discusses all the matching moments. Importance of moment selection is nicely summarized by Hennessy and Whited (2007)[6]. I attempt to match 6 moments: book leverage, CAPM-β, PE ratio, mean earning growth, earning growth volatility and default probability.

[6] "This issue is important since a poor choice of moments can result in large model standard errors in finite samples or an unidentified model. Basing a choice of moments on the size of standard errors constitutes data mining. I choose moments that are a priori informative about parameters. Heuristically, a moment is informative about an unknown parameter if that moment is sensitive to changes in the parameter."

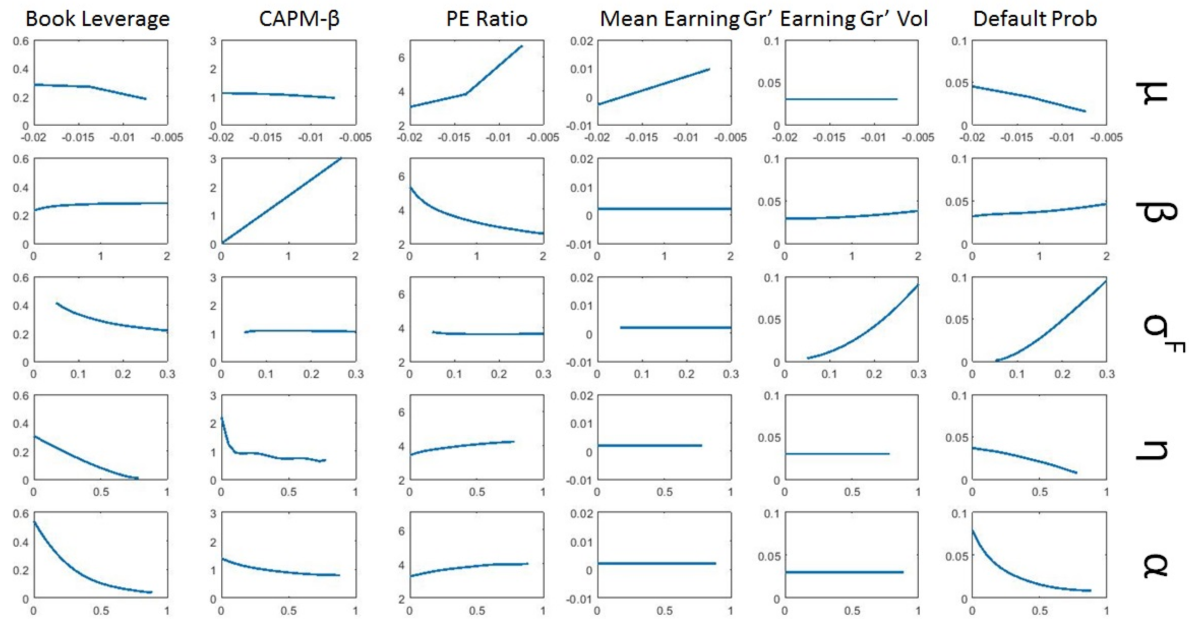

Figure 2: Elasticity of Model Moments with respect to Parameters

Figure 2 illustrates how moments change over parameters and clearly shows which moments help to identify which parameter. As discussed in the last subsection, η and α are identified primarily by the book leverage and default probability. Furthermore, because high η makes equity less risky and thus increases market value of equity (discussed below), it is natural to see that CAPM-β and PE ratio help to identify η.

Now, let us discuss how the remaining three parameters are identified. As expected, μ is pinned down primarily by the earnings growth rate. However, other moments are informative as well. For instance, PE ratio increases in μ. Controlling for the discount rate and aggregate component in the earnings growth rate, a firm with a higher μ has a larger value of equity and thus a higher PE ratio.

Higher β implies higher exposure to the systematic risk. This naturally translates to higher mean CAPM-β. Simultaneously, this implies lower equity price and thus a lower PE ratio. The earning growth rate volatility increase over β and thus helps to identify β. Yet, the earning growth rate volatility better helps to identify its idiosyncratic component (σ_F) than its systematic component (β).

Finally, σ_F is naturally identified by the earning growth rate volatility. Moreover, mean default probability helps to identify σ_F as higher volatility in cash flow increases a probability of reaching the default threshold the next period.

As Figure 2 shows, η is negatively correlated with CAPM-β and this is consistent with empirical findings reported in Garlappi et al. (2008) and Hackbarth et al. (2015). Thus, it is worth discussing how their empirical results relate to the current paper. Their result is based on a model where firms do not internalize higher cost of debt incurred by higher η. As η increases, shareholders expect to recover more upon bankruptcy and thus makes equity less risky (and equity value increases). However, when firms do internalize higher cost of debt, the aforementioned channel is somewhat muted as high η is associated with small default probability. In other words, as default event becomes less likely, the fact that shareholders get to recover more upon bankruptcy matters less. Instead, leverage channel plays a central role in explaining the empirical facts: high η implies low default probability thus makes equity less risky.

5 Data

5.1 Sample Construction

I obtain panel data from CRSP and COMPUSTAT. I align each company's fiscal year appropriately with the calendar year, converting COMPUSTAT fiscal year data to a calendar basis. I inflation-adjust data.[7] I augment it with panel data of corporate marginal tax rates[8]. I impute missing marginal tax rates with time-series average for each firm. Then, I select a sample by deleting firm-quarter observations with missing data. I omit all firms whose primary SIC classification is between 4900 and 4999 or between 6000 and 6999 since the model is inappropriate for regulated or financial firms. Our baseline sample contains 413,689 firm-quarter observations and spans from 1970Q1 to 2016Q4.

[7]I use Consumer Price Index (CPALTT01USQ661S) from OECD https://fred.stlouisfed.org/series/CPALTT01USQ661S

[8]I would like to thank John Graham for sharing panel data of corporate marginal tax rates. https://faculty.fuqua.duke.edu/~jgraham/taxform.html

5.2 Construction of Moments

The paper defines book leverage as $\frac{DLTTQ+DLCQ-CHEQ}{AT}$ where AT, $DLTTQ$, $DLCQ$ and $CHEQ$ are COMPUSTAT codes for total asset, long-term debt, short-term debt and cash. Earning growth is defined as $\tilde{e}_{i,t+1} = \frac{\sum_{j=0}^{K} OIADPQ_{i,t+1-j}}{\sum_{j=0}^{K} OIADPQ_{i,t-j}} - 1$ where K is set to 8. In order to have meaningful earnings growth, I only focus on observations with positive $\sum_{j=0}^{K} OIADPQ_{i,t-j}$. Please note that this still allows both negative and positive earnings growth and simply rules out cases where earning growth's denominator is negative. Similarly, PE ratio is constructed as $\log\left(\frac{\text{PRICE}_{i,t}\cdot\text{Shares}_{i,t}}{(\sum_{j=0}^{K}\text{OIADPQ}_{i,t-j})/K}\right)$. Lastly, I construct CAPM-β based on rolling window of 24 months of monthly returns.

At large, there are two ways to derive default probability. The first is Merton distance to default model, which is based on Merton (1974) bond pricing model. The second is based on Hazard model and is used by a few papers including Campbell et al. (2008). I use the former approach, which is more compatible with the model-implied moments that use Merton-style default probability. Specifically, I follow Bharath and Shumway (2008) to construct default probability[9], which, as Bharath et al. argued, is close to Hazard model's output:

$$\begin{aligned} \pi &= \Phi(-DD) \\ \text{s.t. } DD &= \frac{\ln[(E+F)/F] + (r_{i,t-1} - 0.5\sigma_V^2)T}{\sigma_V\sqrt{T}} \end{aligned}$$

where Φ is a cumulative normal distribution function and σ_V is defined as:

$$\sigma_V = \frac{E}{E+F}\sigma_E + \frac{F}{E+F}(0.05 + 0.25\cdot\sigma_E)$$

Here, σ_E is the annualized percent standard deviation of monthly returns based on trailing 12 months, E is the market value of equity, F is the face value of debt and $r_{i,t-1}$ is annual return calculated by cumulating monthly returns.

5.3 Tax Rates

Following Graham (2000), the literature (e.g. Chen (2010), Glover (2016)) set $\tau_c = 0.35$, $\tau_d = 0.12$ and $\tau_i = 0.296$. However, Graham's sample period covers only from 1980 to

[9] My constructed default probability measures are positively significantly correlated with Moody's commercially available default probability that were used in Garlappi et al. (2008) and Garlappi and Yan (2011). I would like to thank Lorenzo Garlappi for letting me check the correlation.

1994. Because my sample spans from 1970 through 2016, it calls for more up-to-date tax rates. This subsection discusses how tax rates (τ_c, τ_i and τ_d) were constructed.

First, I use corporate marginal tax rate (τ_c) that were constructed according to Graham (1996a) Graham (1996b)[10]. They provide both *before-financing* marginal tax rates (MTR) and *after-financing* MTR. Both measure firm's MTR by incorporating many features present in the tax code, such as tax-loss carryforwards and carrybacks, the investment tax credit, and the alternative minimum tax. *Before-financing* MTR are based on taxable income *before* financing expenses are deducted whereas after-financing MTR are based on taxable income *after* financing expenses are deducted. As Graham (1998) argued, by construction, *after-financing* MTR are endogenously affected by the choice of financing. Because the model treats τ_C exogenous of firms' financing decision, this paper uses *before-financing* MTR.

Second, I closely follow Graham (2000) to construct τ_i and τ_d. As documented in Graham (2000), I set $\tau_i = 47.4\%$ for 1980 and 1981, 40.7% between 1982 and 1986, 33.1% for 1987, 28.7% between 1988 and 1992, and 29.6% afterwards. Based on these estimates for τ_i, I estimate τ_d as $[d+(1-d)g\alpha]\tau_i$. The dividend-payout ratio d is the firm-quarter-specific dividend distribution divided by trailing twelve-quarters moving average of earnings. Since d needs to be less than or equal to 1, if d is greater than 1, I set it to 1. If dividend is missing, I set $d = 0$. The proportion of long-term capital gains that is taxable (g) is 0.4 before 1987 and 1.0 afterwards. I assume that the variable measuring the benefits of deferring capital gains, α, equals 0.25. The long-term capital gains rate, $g\tau_i$ has a maximum value of 0.28 between 1987 and 1997, 0.2 between 1998 and 2003 (Taxpayer Relief Act of 1997) and 0.15 afterwards (Jobs and Growth Tax Relief Reconciliation Act of 2003).

It is worth noting that τ_c is different across firms because firms face different tax-loss carryforwards/carrybacks, the investment tax credit and the alternative minimum tax. τ_d is different across firms because dividend-payout ratios are different. However, for given year, τ_i is the same across firms because I assume that marginal investors face the same τ_i. Also, I assume that τ_c and τ_i stay constant for all four quarters for any given year (due to data limitation) whereas τ_d can potentially change every quarter due to varying dividend-payout ratios.

The above steps yield $\tau_c = 0.2961$ $\tau_i = 0.3318$ $\tau_d = 0.1847$ and $\tau_{cdi} = 0.1038$ on average.

[10] I would like to thank John R. Graham for providing firm-year data for corporate marginal tax rates

Relative to what has been used so far, my τ_c is lower because it captures periods with low earning growth and thus implies lower than statutory tax rates. My τ_i is larger because it accounts for the fact that τ_i is larger in pre-1988 period. Lastly, my τ_d is higher because g is 1 after 1987 and my sample captures more of post-1987 than Graham (2000) does. In net, τ_{cdi} decreased from 0.1320 to 0.1038. As tax shield benefit rates decrease, the Trade-off theory naturally implies lower optimal leverage. As such, more up-to-date tax rates help to partially address underleverage puzzle.

6 Estimation and Results

The objective here is to estimate parameters: μ, β, σ^F, η and α.

6.1 Estimation Procedure

First of all, why do I do simulation at all? Don't I have everything in closed-form solutions? Yes, I do have closed-form functions for firm value, equity value and debt value. But I do not have closed form solutions for matching moments because sample moments are path-dependent, sample is unbalanced panel and the sample suffers from small sample bias. Thus, I need simulations to generate model counterparts.

In order to address firm-specific heterogeneity, I apply firm-fixed effects to the data. More specifically, let us assume that firm i's data at time t is d_{it} (where d is earning growth (eg), book leverage (bl), default probability (dp), CAPM-β ($beta$) or PE ratio (pe)). I construct firm-specific sample average and panel-wide sample average as:

$$\mu_i = \frac{1}{T_i}\sum_{t=1}^{T_i} d_{it}$$

$$\mu = \frac{1}{N}\sum_{i=1}^{N} \mu_i$$

Using this, we convert d_{it} to $\tilde{d}_{it}$ as

$$\tilde{d}_{it} = d_{it} - \mu_i + \mu$$

Then, I construct 6×1 vector M where

$$M = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \left(\sum_{t=1}^{T_i} \begin{bmatrix} \tilde{eg}_{it} \\ \tilde{eg}_{it}^2 \\ \tilde{bl}_{it} \\ \tilde{dp}_{it} \\ \tilde{beta}_{it} \\ \tilde{pe}_{it} \end{bmatrix} \right) = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \left(\sum_{t=1}^{T_i} \begin{bmatrix} eg_{it} \\ \tilde{eg}_{it}^2 \\ bl_{it} \\ dp_{it} \\ beta_{it} \\ pe_{it} \end{bmatrix} \right)$$

Similarly, for parameter θ, for s-th simulated collection of earning sample path, I calculate the model-implied moments $\mathcal{M}_s(\theta)$ as follows:

$$\mathcal{M}_s(\theta) = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \left(\sum_{t=1}^{T_i} \begin{bmatrix} \tilde{eg}_{it,s}(\theta) \\ \tilde{eg}_{it,s}^2(\theta) \\ \tilde{bl}_{it,s}(\theta) \\ \tilde{dp}_{it,s}(\theta) \\ \tilde{beta}_{it,s}(\theta) \\ \tilde{pe}_{it,s}(\theta) \end{bmatrix} \right)$$

Then, I estimate θ by minimizing GMM-weight weighted distance between data moments and model-implied moments:

$$\hat{\theta} = \arg\min_{\theta} \left(M - \frac{1}{S} \sum_{s=1}^{S} \mathcal{M}_s(\theta) \right)' W \left(M - \frac{1}{S} \sum_{s=1}^{S} \mathcal{M}_s(\theta) \right)$$

Here, W is covariance matrix of data-moments after accounting for time-series dependence by clustering by firms (as recommended by Strebulaev and Whited (2012))

$$W = \left(\frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \left[\left(\sum_{t=1}^{T_i} u_{it} \right) \left(\sum_{t=1}^{T_i} u_{it} \right)' \right] \right)^{-1}$$

where u_{it} is an influence function. Since all the moments are sample averages, u_{it} is demeaned value as $u_{it} = \tilde{d}_{it} - m_i$. The standard errors for parameter estimates are given by:

$$\sqrt{N}(\hat{\theta} - \theta_0) \to N \left(0, \left(1 + \frac{1}{S} \right) ((H_0)' W H_0)^{-1} \right)$$

where $H_0 = E\left[\frac{\partial \tilde{M}_s(\theta_0)}{\partial \theta}\right]$. I estimate H_0 by simulating $\tilde{M}_s$ at slightly perturbed $\hat{\theta}$. Lastly, in order to test overidentifications, I define J-statistics as:

$$\left(\sum_{i=1}^{N} T_i\right) \frac{S}{1+S} \left(M - \frac{1}{S}\sum_{s=1}^{S} \mathcal{M}_s(\hat{\theta})\right)' W \left(M - \frac{1}{S}\sum_{s=1}^{S} \mathcal{M}_s(\hat{\theta})\right)$$

I first simulate $S = 100$ time-series of aggregate earning growth. For each time series of aggregate earning growth, I simulate 7922 firm-specific sample path as there are 7922 unique firms in my panel data set. In each simulation, I generate a sample path of $50+T_i$ quarters long cash flow $X_{i,t}$. I discard the first 50 quarters of simulated cash flows to reduce solutions' dependence on $X_{i,t}$ at time $t = 0$. There are 148 quarters (37 years of data) thus, T_i is set to 148.

6.2 Aggregate Parameters

Table 1 summarizes calibrated values for aggregate parameters and the corresponding data sources. In order to match firms' quarterly observation, I use appropriate data counterparts. For aggregate earning growth rate (μ_A) and aggregate earning growth volatility (σ_A), I use quarterly earning series from NIPA. For market Sharpe ratio (φ_A), I use quarterlizied monthly returns from French's website. For risk-free rate (r), I use 3-month risk free rates from CRSP Treasuries. Lastly, I use debt issue cost (ϕ_D) that is reported in Altnklic and Hansen (2000).

	Description	Value	Source
μ_A	Aggr' earning growth	0.017003	NIPA
σ_A	Aggr earning growth vol	0.047857	NIPA
φ_A	Market Sharpe ratio	0.210132	French's data website
r	Risk-free rate	0.0037	CRSP Treasuries
ϕ_D	Prop' debt issuance cost	0.015	Altnklic and Hansen (2000)

Table 1: Aggregate Parameters Value. Calibration Period: 1970Q1-2016Q4

6.3 Results

6.3.1 Full Sample Analysis

In order to test the main null hypothesis H_0, I estimate both restricted model ($\eta = 0$) and unrestricted model. Table 2 shows results for restricted model. As shown, p-value for

χ^2 test is 0.000 and thus I can easily reject the null hypothesis that data are not different from model-implied moments.

	μ	β	σ^F	α	χ^2
estimate	-0.015	0.616	0.187	0.265	156.949
	(0.019)	(0.019)	(0.018)	(0.052)	(0.000)

Table 2: Parameter estimates and standard errors. χ^2 is a chi-squared statistic for the test of the overidentifying restrictions (with p-value in parenthesis)

Table 3 summarizes parameter estimates for unrestricted model. As shown, p-value for χ^2 test is 0.239 and thus I cannot reject the null hypothesis that data are not different from model-implied moments. Most interestingly, η is statistically different from zero thus H_0 can be rejected at 1% significance level. Furthermore, α is consistent with magnitude of bankruptcy cost that the empirical literature found.

	μ	β	σ^F	η	α	χ^2
estimate	-0.015	0.675	0.181	0.070	0.198	1.386
	(0.006)	(0.013)	(0.050)	(0.026)	(0.025)	(0.239)

Table 3: The first five columns show parameter estimates and standard errors in parentheses. χ^2 is a chi-squared statistic for the test of the overidentifying restrictions (with p-value in parenthesis).

In order to compare a restricted model vs. non-restricted model, I perform χ^2 difference test (Newey and West (1987)). If the restriction $\eta = 0$ is true, χ^2 difference should be close to $\chi^2(1)$. However, this is not the case and thus I can simply reject a restricted model ($\eta = 0$) at $p = 0.000$.

Another interpretation of results above is that average firms expect to enter Chapter 11 and expect shareholder to recover non-negative amount upon bankruptcy. If average firms expect to enter Chapter 7 upon bankruptcy, then implied-η should have been 0 but this is statistically significantly rejected. This is consistent with an empirical observation that most of publicly listed firms file for Chapter 11. For example, according to LoPucki bankruptcy database, 97.5% of firms in their sample file for Chapter 11.

Table 4 summarizes mean of data moments (the first column) and model-implied matched moments (the second column). All the moments, especially book leverage and default probability, are matched well. The last column illustrates a counterfactual analysis where the same firm-level estimates as the second column were used except for η where η is set to 0. As shown, setting $\eta = 0$ increased book leverage from 0.276 to 0.311. This illustrates how relaxing $\eta = 0$ constraint partially helps to address underleverage puzzle. It also shows that equity becomes significantly less risky when η increases from 0% set to 7% and this is consistent with Hackbarth et al. (2015)'s empirical finding.

	Data	Simulation	
		Matched Moments	Counterfactual: $\eta = 0$
Book Lev	0.276	0.276	0.311
CAPM-β	1.136	1.140	2.446
PE	3.570	3.556	3.369
Earning Growth	0.005	0.002	0.002
Earning Growth Squared	0.033	0.034	0.034
Default Prob	0.040	0.043	0.045

Table 4: Data and Model-implied Moments.

Now, I discuss which component of the model helps to match empirically observed book leverage: 0.27. Below, I use parameter estimates from Table 3.

	Model 1	Model 2	Model 3	Model 4	Model 5
	static	dynamic	dynamic	dynamic	dynamic
	$\tau_{cdi} = 13.2\%$	$\tau_{cdi} = 13.2\%$	$\tau_{cdi} = 10.38\%$	$\tau_{cdi} = 10.38\%$	$\tau_{cdi} = 10.38\%$
	$\eta = 0$	$\eta = 0$	$\eta = 0$	$\eta = 0.07$	$\eta = 0.07$
Coupon (C)	0.673	0.469	0.372	0.372	0.323
Default Threshold (X_D)	0.201	0.135	0.109	0.117	0.102
Restruc' Boundary (X_U)		2.881	3.304	3.303	3.340
Book Lev	0.468	0.363	0.311	0.304	0.276
1-year Default Prob (%)	5.087	5.286	4.509	4.667	4.367
CAPM-β	2.404	2.498	2.446	1.167	1.139

Table 5: Economic Intuition

Under model 1 (benchmark), I do not allow dynamic restructuring, use tax rates that are typically used in the literature and prevent shareholders from receiving any amount upon bankruptcy. Under this benchmark case, book leverage is 0.468, which is larger than its empirical counterpart.

Next, we allow firms to dynamically restructure (labeled as Model 2). Consistent with the previous literature, I find that allowing upward restructuring allows firms to use debt more conservatively than otherwise: book leverage decrease from 0.468 to 0.363. As noted in Bhamra et al. (2010), shareholders hold a refinancing option that allows them to lever up later. This option makes it possible to reduce the expected costs of financial distress by issuing a smaller amount of debt at each refinancing. Moreover, the presence of the real option to refinance gives rise to yet another difference between the static and dynamic capital structure policies: For the same amount of debt, firms following a dynamic policy postpone default until much later, since the possibility to refinance in the future makes the shareholders' option of waiting to default more valuable. This is noted by smaller X_D decreases from 0.201 to 0.135. Despite lower leverage and lower default threshold, default probability increases from 5.087% to 5.286% mainly due to asymmetry in debt restructuring: firms upward restructure their debt when cash flow is high yet do not downward restructure their debt when cash flow decreases.

Now, I use the most up-to-date tax rates for τ_{cdi} (under Model 3) and study how leverage and default probability change. As tax benefit decreases from 13.2% to 10.38%, obviously firms optimally choose to lower leverage (from 0.363 to 0.311) and thus default probability decreases from 5.286% to 4.509%. This illustrate how using correct tax rates alone can help us to partially address underleverage puzzle.

Under model 4, I allow shareholders to recover non-zero amount upon bankruptcy. However, I force firms to keep the same optimal coupon and restructuring policies as under model 3, i.e. I force firms to use C and X_U that are calculated when $\eta = 0$, in order to study a case when firms do not internalize higher cost due to non-negative η. This clearly shows that default threshold increases and thus default probability increases from 4.509% to 4.667%. Book leverage decreases from 0.311 to 0.304 because book value of equity slightly increases due to non-negative η. Consistent with Hackbarth et al. (2015)'s finding, increase in η significantly decreases CAPM-β from 2.446 to 1.167.

Now, under model 5, finally I allow firms to internalize higher cost of debt and to optimally choose their leverage and restructuring boundary. This further decreases book leverage from 0.304 to 0.276 and default probability from 4.667% to 4.367%. This illustrates how combination of various components of model help us to match a book leverage and default probability. Moreover, this further decreases CAPM-β from 1.167 to 1.139.

6.3.2 Subsample Analysis: Bankruptcy Reform Act 1978

Prior to 1978, as discussed in Section 2, increasing number of firms sought to file under shareholder-friendly Chapter XI[11] (does not require APR to hold) than Chapter X (requires APR to hold). However, that required expensive hearing (see LoPucki and Whitford (1990) for more details). However, Bankruptcy Reform Act 1978 (BRA) completely changed the nature of bankruptcy. Contrary to the prior law, BRA now permitted creditors to take less than full payment, in order to expedite or insure the success of the reorganization. (H.R. Rep No. 595, 95th Cong., 1st Sess. 224 (1978)). This effectively made it easy to deviate from APR.

In fact, Hackbarth et al. (2015) argues that BRA increased η due to four specific clauses. First, relative to the old code, BRA added equity as one additional class to confirm a reorganization plan. Second, managers were given 120-day exclusivity period to propose a plan. Third, if no plan can be agreed upon, a new procedure, called cramdown, allowed firms to continue operating while a buyer was sought. This was considered costly and time-consuming process and thus acted as a disciplinary tool in negotiations in favor of shareholders. Lastly, firms could now declare bankruptcy even when firms were solvent thus debtors can use the threat of bankruptcy as a strategic tool against creditors.

However, not everyone seems to share views with Hackbarth et al. (2015) on BRA's impact. White (1983) suggests that it is not clear whether BRA increased or decreased η. Under the old code (Chapter XI), secured creditors were not allowed to propose or vote for a reorganization plan. But BRA allowed secured-creditors to propose the plan after manager's exclusivity period is over and vote for it. In addition, White suggests that secured creditors were main beneficiary of cramdown procedure because it often obtained a good price and its proceeds are first distributed to secured creditors. White's view was shared by a few articles that were published immediately after BRA 1978 was passed. For example, New York Times 1979 "New Bankruptcy Law: Creditors, Debtors

[11] Consistent with the law literature, Roman letters are used to represent Chandler Act era whereas Arabic letters are used to represent Bankruptcy Reform Act era

Aided" showed that creditors were main beneficiary of BRA. So, it is not clear whether BRA 1978 increased or decreased η. Structural estimation helps me to quantitatively determine BRA's impact.

η is not the only parameter that characterizes bankruptcy process. There is also bankruptcy cost, α, and it is interesting to see how α responded to BRA. Skeel (2001) shows that BRA 1978 reduced the ambiguity present in the bankruptcy law and this could have reduced friction in the bankruptcy process. White (1983) argued that BRA made it harder for badly-run firms to enter reorganization process. Thus, firms that entered reorganization process after 1978 are supposedly better off being reorganized relative to those that entered reorganization process prior to 1978. Thus, BRA could have made the banruptcy process much more efficient by speeding up the asset transfer to more efficient holders and this could have effectively decreased bankruptcy cost α.

In order to test how BRA 1978 impacted η and α, I construct two subsamples: 1970Q1-1978Q3 and 1981Q2-1990Q4. A period between 1978Q4-1981Q1 is removed in accordance with Hackbarth et al. (2015) that claimed the market was still learning of BRA's true impact. First, η is statistically significant at both sample periods. This shows that firms expected APR to be violated upon bankruptcy even before BRA as passed because they expected BRA to eventually pass (the legislative history spanned more than a decade) and firms expected to use shareholder-friendly Chapter XI as opposed to creditor-friendly Chapter X (see Section 2 for some details). Next, η increased from 4% to 5.6% after BRA was passed in 1978. If we form a null hypothesis that η did not change vs. an alternative hypothesis that η increased, the increase is statistically significant at 8% level. More interestingly, BRA had much more significant impact on α: it decreased from 28.4% to 20.2%. Again, if we form a null hypothesis that α did not change vs. an alternative hypothesis that α decreased, the decrease is statistically significant at 2.5% level.

		μ_i	β_i	σ_i^F	η_i	α_i
Sample 1	estimate	-0.032	0.701	0.148	0.040	0.284
1970Q1-1978Q3		(0.015)	(0.000)	(0.019)	(0.006)	(0.034)
Sample 2	estimate	-0.016	0.633	0.182	0.056	0.202
1981Q2-1990Q4		(0.016)	(0.020)	(0.034)	(0.009)	(0.019)
Sample 1 to 2		0.734	-3.482	0.898	1.456	-2.108

Table 6: Time-Series Subsample Analysis. Last row summarize t-test results between Sample 1 and Sample 2

6.3.3 Discussion on magnitude of η

Now, let us compare my subsample-results at Table 6 to the literatures'. Based on sample period from 1979-1986, Weiss (1990) found that APR is violated among 29 cases out of 37 cases. Based on sample period from 1979 to 1986, Eberhart et al. (1990) showed that average shareholder recovery rate is 7.6% upon firms' reorganization. Based on sample period between 1983 and 1990, Franks and Torous (1989) reported shareholders' recovery rate to be 2.28%. Based on sample period between 1982 and 1990, Betker (1995) reported shareholders' recovery rate to be 2.86%. More recently, Bharath et al. (2007) estimated shareholders' recovery rate to be 3.55% between 1979 and 1990,

Literature that are mentioned above estimated shareholders' recovery rate after bankruptcy cost α is realized. Moreover, most of the aforementioned empirical studies looked at post-BRA sample period and thus comparable sample period is Sample 2 (1981-1990). Thus, the comparable number is $\frac{\eta}{1-\alpha} = \frac{0.056}{1-0.202} \approx 7.02\%$. Even though this number sits within the previously documented estimates, it is closer to the upper bound. So, what could explain the discrepancy?

I propose four possible explanations. First, in the spirit of Glover (2016), empirical estimate of η could be biased due to sample selection. Firms with low η have high default probability and those tend to default more frequently than firms with high η. Thus, trying to estimate η by studying only firms that have defaulted could lead researchers to underestimate η. The second source of bias is due to how it was measured. The extant literature typically estimates shareholders' recovery rate by using security prices that most closely postdate the firms' emergence from bankruptcy. However, not every firm successfully emerges from bankruptcy. Thus, studying η only among firms that have successfully emerged from bankruptcy could potentially bias η's estimate.

Third, even though η captures shareholders' recovery rate upon bankruptcy in the model, it is meant to capture all the benefits that shareholders have, which the current model abstracts away from, such as benefit of control rights as discussed in Section 3.4. Thus, my estimate of η could be larger than what is actually observed among small set of bankrupt firms. Lastly, the literature's empirical estimate for η could be incorrect mainly due to how the market value of equity and warrant is priced. If shareholders do receive non-zero value upon bankruptcy, they typically receive new firms' stocks and warrants. Thus, it is important to correctly estimate market value of these two instruments and incorrect market value certainly implies incorrect measure for η. I can illustrate this in Wickes' bankruptcy case. Using its first publicly available stock price after emergence (Jan, 1985), Eberhart et al. (1990) estimated shareholders' recovery rate as 4%. However, there is no reason to use this particular date's price over price on other dates. For example, if average shareholders kept their shares for 17 months, price on May, 1986 could be more appropriate to use and η would be 6%, 50% larger than Eberhart et al.'s estimate. Thus, η that were documented in the literature could be inaccurate.

6.3.4 Subsample Analysis: Proxies for η

The results so far quantify how much shareholders expect to recover upon bankruptcy for the average firm. Now, I begin exploring how these values vary over firms with different characteristics. Several papers attempted to empirically use η in various contexts. However, because it is unobservable, they had to rely on observable proxies. Due to lack of guidance on proxies' validity, the literature used wide range of different measures. Unfortunately, in many cases, these empirical measures simultaneously proxy other unobservable firm characteristics and thus its validities are ambiguous as admitted by Davydenko and Strebulaev (2007). This subsection first lists a few commonly-used proxies. Then, I conduct subsample analysis to test their validity.

First, citing more frequent violation[12] of APR in favor of shareholders in larger firms, Garlappi et al. (2008) Garlappi and Yan (2011) and Hackbarth et al. (2015) used firm size as a positive proxy for η. They argue that small firms usually have higher concentration

[12]Please see Weiss (1990) Betker (1995) and Franks and Torous (1994)

of bond ownership and thus during financial distress, this concentration of, and close monitoring by, creditors severely weakens η. However, perhaps one can also argue that small firms' shareholders tend to be more concentrated and thus this could cancel out a force that bond ownership concentration has. Thus, it is unclear how well firm size proxies η. I use log of total asset ot measure firm size.

Second, Hackbarth et al. (2015) and Garlappi et al. (2008) used tangibility as a negative proxy for η. Intuitively, creditors of firms with more tangible assets find it easier and more profitable to liquidate. Thus, creditors are less incentivised to negotiate with shareholders thus leads to lower η. However, tangibility is widely used to proxy α as well, thus it is unclear whether tangibility can be a good proxy for both α and η. What measure do I use? According to Berger et al. (1996), 1 dollar of book asset value generates on average 71.5 cents in exit value for total receivables, 54.7 cents for inventory and 53.5 cents for capital. Accordingly, Garlappi et al. (2008) defined tangibility as $0.715 \cdot \text{Receivables}(RECTQ) + 0.547 \cdot \text{Inventory}(INVTQ) + 0.535 \cdot \text{Capital}(PPENTQ)$. I believe that this tangibility measure better captures η and α because tangibility is relevant only when it can be sold at certain price upon bankruptcy. Thus, Berger's measure is a better proxy than other alternative measures (e.g. $PPEGT/AT$) that capture the gross level of tangible asset.

Third, various papers have used intangibility measure as a proxy for η. Yet, its implied sign is questionable. Tangibility can potentially be negatively correlated with intangibility measure. If so, the above argument that tangibility is a negative proxy for η should imply that intangibility is a *positive* proxy. Yet, Garlappi et al. (2008) Garlappi and Yan (2011) and Hackbarth et al. (2015) used it as a *negative* proxy. Firms with high intangibility are more likely to face liquidity shortage (Opler and Titman (1994)) during financial crisis thus are more likely to forgo intangible investment opportunities that shareholders value (Lyandres and Zhdanov (2013)). Firms' urgent need for liquidity effectively acts as cash-flow based covenants and thus high intangibility puts shareholders at disadvantage vs. creditors and implies low η. I use an intangible measure that aggregates all the investment in intangible asset over years (proposed by Peters and Taylor (2016)).

Then, I follow standard procedure to form 3 buckets based on each proxy. For each proxy, I restrict analysis to panel data set with non-missing data for proxies. Then, I estimate η and α for each bucket independently. Table 7 shows estimates for η at the upper panel and estimates for α at the lower panel.

	1	2	3
Estimates for η			
Size	0.062	0.071	0.074
	(0.017)	(0.086)	(0.008)
Tangibility	0.088	0.067	0.056
	(0.101)	(0.019)	(0.024)
Intangibility	0.076	0.067	0.067
	(0.017)	(0.015)	(0.038)
Estimate for α			
Size	0.186	0.193	0.235
	(0.049)	(0.049)	(0.018)
Tangibility	0.192	0.224	0.202
	(0.113)	(0.009)	(0.035)
Intangibility	0.163	0.225	0.228
	(0.018)	(0.029)	(0.023)

Table 7: Upper panel: estimates for η with standard errors in parenthesis. Lower panel: estimates for α with standard errors in parenthesis

First, let us look at firm size results. η increases almost monotonically as firm size increases. This can partially explain why underleverage is pronounced among larger firms. I would like to note that bankruptcy cost α increases over firm size. This pattern could partially alleviate a concern that larger η might be mainly driven by smaller α because if it were, α should have decreased over firm size. Next, let us observe tangibility results: η decreases over tangibility. Thus, the literature's practice to use tangibility as a negative proxy for η seems to be valid. There is no obvious pattern on how α changes over tangibility. Lastly, let us study intangibility results. Consistent with our intuition, α increases and η decreases as firms contain more intangible asset.

7 Robustness Check: Firm-Level Estimation

Limitation of the structural estimation (discussed in Section 6.1, referred as "W" to represent that it uses the whole sample) is that it assumes that firms are homogeneous.

However, more accurate quantitative analysis could be done if I can address heterogeneity in firms. To that end, I follow Glover (2016)'s non-parametric estimation method at firm-level (Please see Section B for more details). Throughout this section, I refer to this estimation method as "F" to represent that it uses an individual firm-level time-series data.

F could help us to confirm our results in the previous section and test validity of empirical proxies for η. Moreover, sign of cross-sectional correlation between α and η sheds some light on validity of renegotiation model first introduced in Fan and Sundaresan (2000). However, despite the richness in results that F can possibly bring, F estimates could suffer from bias due to two sources: Jensen's inequality bias and sample selection bias. Section 7.1 studies how this bias changes over two estimation procedures, W and F when different data generating process assumptions are made. Noting these caveats, I cautiously report results in Section 7.2.

7.1 Bias in Firm-level estimation

7.1.1 Jensen's Inequality Bias

Jensen's inequality bias arises due to two reasons: 1) main matching moment, leverage, is a convex function of main parameters, α and η and 2) firm heterogeneity. In this subsection, I explain these two reasons in detail and how that contributes to bias.

First, as shown in Equation (3), leverage is a convex function of α and η. Because α and η act as a commitment device, high α and high η imply lower optimal leverage. For illustration, let us think of firm 1 whose α is smaller than firm 2's. Then, firm 1's leverage is larger than firm 2's. The economic model implies that the marginal benefit of further increase in debt declines as debt increases. Thus, firm 1's marginal benefit of further increase in debt is smaller than firm 2's. This means that elasticity of firm 2's optimal leverage with respect to α is lower than firm 1's and thus explains a convexity.[13]. More technical explanation is as follows. When $\alpha = 0$, debt is almost risk-free and thus optimal leverage is 100%. As α approaches 1, debt becomes very risky and thus optimal

[13] The fact that firm 1's leverage is larger than firm 2's could imply that firm 1's marginal cost is larger than firm 2's. Yet, because marginal cost is also positively correlated with bankruptcy cost α, these two effects roughly cancel each other out and thus firm 1 and firm 2 have similar marginal cost.

leverage approaches 0 but is bounded below at 0. Thus, leverage should decrease at smaller absolute rate over α as α increases. In other words, leverage should be a convex function of α. The similar intuition applies to η.

Second, heterogeneity can arise due to two sources: heterogeneous parameter values or heterogeneous omitted variables. For this, I use a simulation. Let us think of a hypothetical convex function:

$$f(\kappa) = \frac{1}{800}\kappa^{-2}$$

where κ can correspond to α or η in my model. Let us assume that the observable leverage is

$$f(\kappa) + \epsilon$$

where ϵ captures omitted variable. Now, I sample κ and ϵ N-times from the joint-normal distribution:

$$\begin{bmatrix} \kappa \\ \epsilon \end{bmatrix} \sim N\left(\begin{pmatrix} 0.2 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_\kappa^2 & \rho\sigma_\kappa\sigma_\epsilon \\ \rho\sigma_\kappa\sigma_\epsilon & \sigma_\epsilon^2 \end{bmatrix} \right)$$

where ρ is a correlation between κ and ϵ. When $\rho \neq 0$, this captures an endogeneity issue due to an omitted variable. For homogeneous parameter value case, I set $\sigma_\kappa = 0$. For heterogeneous parameter value case, I set $\sigma_\kappa = 0.02$. For no omitted variable case, I set $\sigma_\epsilon = 0$. For medium omitted variable case, I set $\sigma_\epsilon = 0.02$. For large omitted variable case, I set $\sigma_\epsilon = 0.04$. Here, N represents the number of observation in the sample.

As summarized in Table 8, both estimations suffer from small sample bias, especially when it is compounded with heterogeneous parameters and/or omitted variables. However, it is interesting to see that whole-sample estimates is always biased downward and bias is relatively small when N is sufficiently large. Yet, the firm-level estimates' bias is very significant even when N is sufficiently large. The performance difference between W and F is mainly due to Jensen's inequality.

		$N=2$		$N=100$		$N=500$		$N=2000$	
		W	F	W	F	W	F	W	F
				ρ is -0.5.					
homo	no omitted var	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200
	med	0.185	0.206	0.202	0.294	0.198	0.282	0.196	0.274
	large	0.173	0.306	0.189	0.393	0.185	0.384	0.182	0.379
hetero	no omitted var	0.203	0.205	0.196	0.199	0.197	0.200	0.197	0.201
	med	0.293	0.604	0.193	0.297	0.196	0.305	0.196	0.305
	large	0.333	0.618	0.180	0.412	0.184	0.405	0.183	0.409
				ρ is 0.5.					
homo	no omitted var	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200
	med	0.185	0.206	0.202	0.294	0.198	0.282	0.196	0.274
	large	0.173	0.306	0.189	0.393	0.185	0.384	0.182	0.379
hetero	no omitted var	0.203	0.205	0.196	0.199	0.197	0.200	0.197	0.201
	med	0.272	0.273	0.199	0.275	0.198	0.259	0.196	0.254
	large	0.493	0.675	0.189	0.355	0.188	0.352	0.185	0.368

Table 8: Comparison between whole-sample estimation ("W") vs. firm-level estimation ("F"). Homogeneous case captures $\sigma_\kappa = 0$ and heterogeneous case captures $\sigma_\kappa = 0.02$. For no omitted variable case, I set $\sigma_\epsilon = 0$. I set $\sigma_\epsilon = 0.02$ and $\sigma_\epsilon = 0.04$ respectively for medium and large level of omitted variable case. Lastly, N represents the sample size.

Let us first discuss W. W's estimation procedure gets us

$$f^{-1}(E_N(f(\kappa)+\epsilon)) = f^{-1}(E_N(f(\kappa)) + E_N(\epsilon))$$

where E_N is a notation for sample average. When N is small, it suffers from small sample bias as $E_N(\epsilon)$ is not necessarily equal to 0. Thus, W estimates can vary across different level of omitted variable problems. However, when N is sufficiently large, $E_N(\epsilon) \approx 0$ and thus W estimate gets close to $f^{-1}(E_N(f(\kappa)))$ Thus, omitted variable ϵ does not play significant role and explains why W estimates are not sensitive to level of omitted variables nor level of endogeneity (ρ). Lastly, it is interesting to see that W estimates do not suffer from parameter heterogeneity issue when we have sufficiently large sample size.

Now, let us discuss F. F's estimation procedure gets us $E_N[f^{-1}(f(\alpha)+\epsilon)]$. Thus, even when N is sufficiently large, omitted variable plays a significant role and explains why F estimates are very sensitive to the level of omitted variables and level of endogeneity.

The limitation of F estimation procedure can potentially explain why Glover (2016) estimated bankruptcy cost to be significantly high at 45%, which is at least twice as large as what is empirically observed. This simulates and confirms Reindl et al. (2017)'s critique that Glover's estimate could suffer from omitted variables. Also, by showing that W estimation does not significantly suffer from heterogeneous parameter or omitted variables problems, my main estimation results in Section 6 are somewhat robust. Section C presents more intuition on how W and F fare by illustrating a simple example.

Finally, I investigate a potential bias in cross-sectional regression. Here, I only study a case where true κ is heterogeneous where a cross-sectional regression is meaningful. Then, I regress estimated $\hat{\kappa}$ on true κ to see how regression bias varies over different degree of omitted variables and endogeneity.

	$N = 2$	$N = 100$	$N = 500$	$N = 2000$
Correlation between α and ϵ is -0.5.				
no omitted var	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
med	3.088 (10.731)	1.558 (0.119)	1.579 (0.054)	1.572 (0.027)
large	3.151 (10.241)	2.147 (0.178)	2.101 (0.075)	2.112 (0.038)
Correlation between α and ϵ is 0.5.				
no omitted var	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
med	1.314 (0.948)	1.356 (0.114)	1.263 (0.045)	1.240 (0.021)
large	3.123 (12.060)	1.732 (0.170)	1.699 (0.074)	1.767 (0.038)

Table 9: Regression coefficient: $\hat{\alpha}$ on true α for heterogeneous case

It is assuring to see that small sample bias decreases as N increases. For all N, if omitted variable does not exist, then regression coefficient is not biased. However, as the degree of omitted variable increases, a bias in the regression coefficient increases. Thus, regression results reported in Section 7.2 could potentially suffer from bias due to omitted variables.

7.1.2 Data and Sample Selection Bias

In order to make sure that I have enough time-series data, I limit sample to firms with at least 20 quarters of observations. Second, I remove firm-quarter observations with negative tax benefit, i.e. negative τ_{cdi}. Only reason firms lever up in the model is to gain tax benefit and thus the model is not appropriate to analyze firms with negative τ_{cdi}. After applying criteria mentioned above, the sample contains contain 2,804 unique firms.

Fortunately, firm-level estimation allows me to use different marginal tax rates for different firms. Table 10 summarizes τ_c, τ_i, τ_d and tax shield benefit (τ_{cdi}). As shown, tax rates, especially tax shield benefits, are widely dispersed across firms and time. Thus, the usual practice to impose one-size-fits-all tax rate might not be applicable, especially in studying trade-off theory.

	τ_c	τ_i	τ_d	τ_{cdi}
Mean	0.348	0.343	0.200	0.136
Median	0.350	0.322	0.196	0.125
Std Dev	0.063	0.055	0.129	0.075

Table 10: Summary Statistics for Tax for 2,804 firms

Unfortunately, this sample selection practice introduces sample selection bias. For example, firms with at least 20 quarters observations tend to make higher earning and thus have high interest tax benefits (13.6%) than average firms with low earning (9.64%). Nonetheless, I believe that firm-level estimates provide additional insights that the aggregate estimates do not provide and thus I report results below.

7.2 Results

Despite this limitation, these results could be informative as a robustness check and thus I report those below.

7.2.1 Parameter Estimates

Table 11 summarizes parameter estimates. The first row summarizes the cross-sectional mean of firm-level parameter estimates, the second row shows the cross-sectional mean of firm-level parameter standard error and the third row shows the cross-sectional median of firm-level parameters. Most relevantly, η is still statistically different from zero thus H_0 can be rejected at 0.1% significance level.

	μ	β	σ^F	η	α
mean	-0.017	0.723	0.110	0.104	0.388
	(0.008)	(0.033)	(0.014)	(0.028)	(0.035)
median	-0.016	0.679	0.096	0.089	0.318

Table 11: The first row show mean of firm-level parameter estimates, the second row shows mean of firm-level parameter standard errors in parentheses and the third row shows median of firm-level parameter estimates.

Table 12 summarizes cross-sectional correlation among firm-level parameters. Correlation is:

	μ	β	σ^F	η
β	0.340			
σ^F	-0.078	-0.010		
η	-0.122	-0.030	-0.105	
α	-0.030	-0.037	-0.174	-0.071

Table 12: Correlation of Parameters

It is interesting to note that α and η are negatively correlated. If true, this could cast doubt on how Fan and Sundaresan (2000) modeled. They used bankruptcy cost α as a bargaining surplus in modeling the renegotiation between shareholders and creditors upon bankruptcy. Thus, in their model, shareholders recover a fraction of α upon bankruptcy and thus implies that η and α should be positively correlated. This inconsistency needs to be revisited.

7.2.2 Bias in the Estimates

In this subsection, I compare different structural estimation results in order to highlight Jensen's Inequality bias and Sample Selection bias and how these comparisons still hold even when we allow APR to be relaxed. I summarize results in Table 13.

	Whole Sample (7,922)	Firm-level Sample (2,804)
W	$\alpha = 26.5\%, \eta = 0\%$	$\alpha = 29\%, \eta = 0\%$
	$\alpha = 19.8\%, \eta = 7\%$	$\alpha = 22.5\%, \eta = 8\%$
F		$\alpha = 45\%, \eta = 0\%$
		$\alpha = 38.8\%, \eta = 10.4\%$

Table 13: Estimation Results for different sample using different estimation methods

The table shows 6 structural estimation results in total. Two structural estimation results in W-row and Whole-sample column reproduces results from Table 2 and 3 where the first line shows a restricted model's result (APR is imposed) and the second line shows non-restricted model's result (APR is relaxed). In this subsection, I only discuss restricted model's result. But the similar argument still holds for both α and η when we relax APR.

Let us first test how sample selection bias (discussed in Section 7.1.2) alters estimates by looking at two structural estimation results in W-row and Firm-level sample column. Restricting data to firms with at least 20 quarters time-series and positive tax benefits selects firms with certain characteristics such as low earning volatility. Because these sample restriction *did not* change book leverage and lower σ^F implies higher book leverage, I need much larger α to match the book leverage. Thus, in order to match the same book leverage with smaller earning growth volatility, α has to be larger by 3 percentage points.

Now, let us observe how Jensen's inequality bias (discussed in Section 7.1.1) changes estimate by looking at two structural estimation results in F-row and Firm-level sample. These results are consistent with Glover (2016)'s and my results in Table 11. When we use firm-level sample, F estimate of α is significantly larger than W estimate of α. This is consistent with my simulation results when estimates suffer from omitted variables problems in Table 8. Thus, this illustrates that Glover (2016)'s large estimate of α could be due to omitted variable problems.

7.2.3 η and empirical proxies

This subsection revisits Section 6.3.4 and studies validity of empirical proxies by looking at firm-level η. I regress firm-level η's on empirical proxies and summarize results in Table 14.

	(1)	(2)	(3)	(4)
Firm Size	0.003			0.004
	(0.001)			(0.001)
Tangibility		0.044		0.070
		(0.016)		(0.017)
Intangibility			-0.005	0.003
			(0.004)	(0.005)
N	2,804	2,804	2,804	2,804
Adjusted-R^2(%)	0.372	0.231	0.010	0.871

Table 14: Regression estimates of η on empirical proxies.

Consistent with Section 6.3.4, firm size is statistically significant positive proxy for η. Regression estimate does not change much even after we control for other proxies as we look at column (1) and (4). As firms increase in size by a factor of 2, η increases by $0.3\log(2) = 0.21\%$

However, results for other proxies are not always consistent with Section 6.3.4's. Whereas subsample analysis yields that tangibility is a negative proxy for η, firm-level estimates imply that tangibility is a positive proxy for η. Second, even though subsample analysis yield that intangibility is a negative proxies for η, my firm-level estimates show that those results are not statistically significant. At best, these imply that we need to rethink about using tangibility and intangibility as proxies for η.

8 Conclusion

I address underleverage puzzle by allowing APR to be violated. To that end, I structurally estimate shareholders' expected recovery rate. Shareholders' strategic default action, whose severity is determined by shareholders' recovery rate, acts as a "negative" commitment device. Thus, firms' optimal leverage decreases over shareholders' recovery rate. This channel helps to match empirically observed leverage and default probability. Structural estimation yields 19.8% of expected bankruptcy cost and 7% of shareholders' expected recovery rate, both of which are in line with the previous literature' finding. Time-series subsample analysis reveals that Bankruptcy Reform Act 1978 increased shareholders' expected recovery rate and decreased bankruptcy cost. Furthermore, consistent with the empirical literature, my subsample and firm-level estimation results show that firm size is a good positive proxy for shareholders' expected recovery rate and can potentially explain why underleverage puzzle seems to be pronounced among large firms.

Lastly, even though my paper attempts to answer a positive question, this framework can be used to shed some light on an important policy question: what is the optimal bankruptcy procedure? By giving the bankrupt firms the second chance to redeem themselves, shareholder-friendly bankruptcy policy is certainly a popular policy. However, it will make debt costly, which will eventually lead to debt reduction and potentially retard the economic development. My paper highlights its unintended consequences.

References

H. Almedia and T. Philippon. The risk-adjusted coste of financial distress. *Journal of Finance*, 62:2557–2586, 2007.

E. Altman. A further empirical investigation of the bankruptcy cost question. *Journal of Finance*, 39:1067–1089, 1984.

O. Altnklic and R. S. Hansen. Are there economies of scale in underwriting fees? evidence of rising external financing costs. *Review of Financial Studies*, 13:191–218, 2000.

G. Andrade and S. N. Kaplan. How costly is financial (not economic) distress? evidence from highly leveraged transactions that became distressed. *Journal of Finance*, 53: 1443–1493, 1998.

P. Berger, E. Ofek, and I. Swary. Investor valuation of the abandonment option. *Journal of Financial Economics*, 42:257–287, 1996.

B. Betker. Management's incentive, equity bargaining power, and deviation from absolute priority. *Journal of Business*, 68:161–184, 1995.

H. S. Bhamra, L.-A. Kuehn, and I. A. Strebulaev. The aggregate dynamics of capital structure and macroeconomic risk. *Review of Financial Studies*, 23:4187–4241, 2010.

S. T. Bharath and T. Shumway. Forecasting default with the merton distance to default model. *Journal of Financial Economics*, 21:1339–1369, 2008.

S. T. Bharath, V. Panchapegesan, and I. Werner. The changing nature of chapter 11. *Working Paper*, 2007.

Y. Boualam, J. F. Gomes, and C. Ward. Understanding the behavior of distressed stocks. *Working Paper*, 2017.

J. Y. Campbell, J. Hilscher, and J. Szilagyi. In search of distress risk. *Journal of Finance*, 63:2899–2939, 2008.

H. Chen. Macroeconomic conditions and the puzzles of credit spreads and capital structure. *Journal of Finance*, 65:2171–2212, 2010.

S. Davydenko and I. Strebulaev. Strategic actions and credit spreads: An empirical investigation. *Journal of Finance*, 62:2633–2671, 2007.

S. A. Davydenko and J. R. Franks. Do bankruptcy codes matter? a study of defaults in france, germany, and the u.k. *Journal of Finance*, 63:565–608, 2008.

S. A. Davydenko, I. A. Strebulaev, and X. Zhao. A market-based study of the cost of default. *Review of Financial Studies*, 25:2959–2999, 2012.

D. Duffie and K. J. Singleton. Simulated moments estimation of markov models of asset prices. *Econometrica*, 61:929–952, 1993.

A. C. Eberhart, W. T. Moore, and R. L. Roenfeldt. Security pricing and deviations from the absolute priority rule in bankruptcy proceedings. *Journal of Finance*, 45:1457–1469, 1990.

R. Elkamhi, J. Ericsson, and C. A. Parsons. The cost and timing of financial distress. *Journal of Fincial Economics*, 105:62–81, 2012.

H. Fan and S. Sundaresan. Debt valuation, renegotiation, and optimal dividend policy. *Review of Financial Studies*, 13:1057–1099, 2000.

J. R. Franks and W. N. Torous. An empirical investigation of u.s. firms in renegotiation. *Journal of Finance*, 44:747–779, 1989.

J. R. Franks and W. N. Torous. A comparison of financial recontracting in distressed exchanges and chapter 11 reorganizations. *Journal of Financial Economics*, 35:349–370, 1994.

L. Garlappi and H. Yan. Financial distress and the cross-section of equity returns. *Journal of Finance*, 66:789–822, 2011.

L. Garlappi, T. Shu, and H. Yan. Default risk, shareholder advantage, and stock returns. *Review of Financial Studies*, 21:2743–2778, 2008.

B. Glover. The expected cost of default. *Journal of Financial Economics*, 119:284–299, 2016.

R. Goldstein, N. Ju, and H. Leland. An ebit-based model of dynamic capital structure. *Journal of Business*, 74:483–512, 2001.

J. Graham. Do personal taxes affect corporate financing decision? *Journal of Public Economics*, 73:147–185, 1999.

J. Graham. Taxes and corporate finance: A review. *Review of Financial Studies*, 16: 1075–1129, 2003.

J. R. Graham. Debt and the marginal tax rate. *Journal of Financial Economics*, 41: 41–73, 1996a.

J. R. Graham. Proxies for the corporate marginal tax rate. *Journal of Financial Economics*, 42:187–221, 1996b.

J. R. Graham. Debt, leases, taxes, and the endogeneity of corporate tax status. *Journal of Finance*, 53:131–162, 1998.

J. R. Graham. How big are the tax benefits of debt? *Journal of Finance*, 55:1901–1941, 2000.

J. R. Graham, M. T. Leary, and M. R. Roberts. A century of capital structure: The leveraging of corporate america. *Journal of Financial Economics*, 118:658–683, 2015.

D. Green. Corporate refinancing, covenants and the agency cost of debt. *Working Paper*, 2018.

D. Hackbarth, R. Haselmann, and D. Schoenherr. Financial distress, stock returns, and the 1978 bankruptcy reform acts. *Review of Financial Studies*, 28:1810–1847, 2015.

J. M. Harrison. *Brownian Motion and Stochastic Flow Systems*. John Wiley and Sons, New York, 1985.

C. Hennessy and T. Whited. How costly is external financing? evidence from a structural estimation. *Journal of Finance*, 62:1705–1745, 2007.

N. Ju, R. Parrino, A. M. Poteshman, and M. S. Weisbach. Horses and rabbits? trade-off theory and optimal capital structure. *Journal of Financial and Quantitative Analysis*, 40:259–281, 2005.

L. P. King. Chapter 11 of the 1978 bankruptcy code. *American Bankruptcy Law Journal*, 53:107–132, 1979.

H. Leland. Corporate debt value, bond covenants and optimal capital structure. *Journal of Finance*, 49:1213–52, 1994.

M. L. Lemmon and J. F. Zender. Looking under the lamppost: An empirical examination of the determinants of capital structure. *Working Paper*, 2001.

L. M. LoPucki and W. C. Whitford. Bargaining over equity's share in the bankruptcy reorganization of large, publicly held companies. *University of Pennsylvania Law Review*, 139:125–196, 1990.

E. Lyandres and A. Zhdanov. Investment opportunities and bankruptcy prediction. *Journal of Financial Markets*, 16:439–476, 2013.

R. C. Merton. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29:449–470, 1974.

M. H. Miller. Debt and taxes. *Journal of Finance*, 32:261–275, 1977.

W. K. Newey and K. D. West. Hypothesis testing with efficient method of moments estimation. *Internatoinal Economic Review*, 28:777–787, 1987.

T. Opler and S. Titman. Financial distress and corporate performance. *Journal of Finance*, 49:1015–1040, 1994.

R. H. Peters and L. A. Taylor. Intangible capital and the investment-q relation. *Journal of Financial Economics*, Forthcoming, 2016.

E. A. Posner. The political economy of the bankruptcy reform act of 1978. *Michigan Law Review*, 96:47–126, 1997.

J. Reindl, N. Stoughton, and J. Zechner. Market implied costs of bankruptcy. *Working Paper*, 2017.

D. Skeel. Creditors' ball: The 'new' new corporate governance in chapter 11. *Faculty Scholarship. Paper 29*, 2003.

D. A. Skeel. *Debt's Dominion: A History of Bankruptcy Law in America.* Princeton University Press, 2001.

I. A. Strebulaev and T. M. Whited. *Dynamic Models and Structural Estimation in Corporate Finance.* Now Publishers Inc, 2012.

J. H. van Binsbergen, J. R. Graham, and J. Yang. The cost of debt. *Journal of Finance*, 65:2089–2136, 2010.

L. A. Weiss. Bankruptcy resolution: Direct costs and violation of priority of claims. *Journal of Financial Economics*, 27:285–314, 1990.

L. A. Weiss and K. H. Wruck. Information problems, conflicts of interest and asset stripping: Chapter 11's failure in the case of eastern airlines. *Journal of Financial Economics*, 48:55–97, 1998.

M. J. White. Bankruptcy costs and the new bankruptcy code. *Journal of Finance*, 38: 477–488, 1983.

A Smooth Pasting Condition

As a reminder, a function for equity value is

$$E(X_t) = \frac{1-\tau_{cd}}{r-\hat{\mu}}X_t - \frac{(1-\tau_{cd})C}{r} + B_1 X_t^{\lambda_+} + B_2 X_t^{\lambda_-}$$

First, because X_D is chosen to maximize $E(X)$, we need to have:

$$B_1'(X_D) = 0 \text{ and } B_2'(X_D) = 0$$

Second, value matching condition specifies that

$$\frac{1-\tau_{cd}}{r-\hat{\mu}}X_D - \frac{(1-\tau_{cd})C}{r} + B_1(X_D)X_D^{\lambda_+} + B_2(X_D)X_D^{\lambda_-} = \frac{\eta(1-\tau_{cd})X_D}{r-\hat{\mu}}$$

where B_1 and B_2 are functions of X_D. Let us take a derivative of both sides with respect to X_D

$$\frac{1-\tau_{cd}}{r-\hat{\mu}} + B_1'(X_D)X_D^{\lambda_+} + B_1(X_D)\lambda_+ X_D^{\lambda_+-1} + B_2'(X_D)X_D^{\lambda_-} + B_2(X_D)\lambda_- X_D^{\lambda_--1} = \frac{\eta(1-\tau_{cd})}{r-\hat{\mu}}$$

Substituting $B_1'(X_D) = 0$ and $B_2'(X_D) = 0$, we have:

$$\frac{1-\tau_{cd}}{r-\hat{\mu}} + B_1(X_D)\lambda_+ X_D^{\lambda_+-1} + B_2(X_D)\lambda_- X_D^{\lambda_--1} = \frac{\eta(1-\tau_{cd})}{r-\hat{\mu}}$$

Thus, we have:

$$\lim_{X_t \downarrow X_D} E^{'}(X_t) = \frac{\eta(1-\tau_{cd})}{r-\hat{\mu}}$$

B Firm-Level Estimation Procedure

Firm-level estimation procedure takes advantage of each firm's time-series variation. Similar to whole-sample estimation procedure, I match 6 moments in order to estimate 5 parameters. To avoid confusion, in this subsection, I explicitly put subscript i for firm i. For each firm i, I use SMM to estimate firm-specific parameter values: $\theta_i = [\mu_i,\ \beta_i,\ \sigma_i^F,\ \eta_i,\ \alpha_i]$. For each moment, I define data-moment M^i as follows:

$$M^i = \frac{1}{T_i}\sum_{t=1}^{T_i}\begin{bmatrix} eg_{it} \\ eg_{it}^2 \\ bl_{it} \\ dp_{it} \\ beta_{it} \\ pe_{it} \end{bmatrix}$$

I define simulated-moment $\mathcal{M}_s^i(\theta_i)$ as:

$$\mathcal{M}_s^i(\theta_i) = \frac{1}{T_i}\sum_{t=1}^{T_i}\begin{bmatrix} eg_{it,s}(\theta_i) \\ eg_{it,s}^2(\theta_i) \\ bl_{it,s}(\theta_i) \\ dp_{it,s}(\theta_i) \\ beta_{it,s}(\theta_i) \\ pe_{it,s}(\theta_i) \end{bmatrix}$$

$d_{it,s}(\theta_i)$ is the simulated observation at date t for firm i. Then, I estimate θ_i as

$$\hat{\theta}_i = \arg\min_{\theta} \left(M^i - \frac{1}{S}\sum_{s=1}^{S} \mathcal{M}_s^i(\theta) \right)' W^i \left(M^i - \frac{1}{S}\sum_{s=1}^{S} \mathcal{M}_s^i(\theta) \right)$$

where

$$W^i = \left(\sum_{j=-k}^{k} \left(\frac{k-|j|}{k} \right) \frac{1}{T_i} \sum_{t=1}^{T_i} (u_{i,t} u_{i,t-j}^{'}) \right)^{-1}$$

where $u_{i,t} = d_{i,t} - \left(\frac{1}{T_i} \sum_{t=1}^{T_i} d_{i,t} \right)$.

Once θ_i is estimated, I calculate its standard error. As shown by Duffie and Singleton (1993), the distribution of $\hat{\theta}_i$ is

$$\sqrt{T_i}(\hat{\theta}_i - \theta_{0,i}) \to N \left(0, \left(1 + \frac{1}{S} \right) ((H_0^i)' W^i (H_0^i))^{-1} \right)$$

where $H_0^i = E\left[\frac{\partial \tilde{M}_s^i(\theta_{0,i})}{\partial \theta} \right]$. I estimate H_0^i by simulating $\tilde{M}_s^i$ at slightly perturbed $\hat{\theta}_i$. Then, I calculate standard errors for $\hat{\theta}_i$. Lastly, in order to test overidentifications, I define J-statistics as:

$$\frac{T_i S}{1+S} \left(M^i - \frac{1}{S}\sum_{s=1}^{S} \tilde{M}_s^i(\hat{\theta}_i) \right)' W^i \left(M^i - \frac{1}{S}\sum_{s=1}^{S} \tilde{M}_s^i(\hat{\theta}_i) \right)$$

I first simulate 50 time-series of aggregate earning growth. For each time series of aggregate earning growth, I simulate 100 firm-specific sample path and thus $S = 50 \times 50 = 2500$. I do this for each of firms in my sample[14]. In each simulation, I generate a sample path of 400+T_i quarters long cash flow $X_{i,t}$. I discard the first 400 quarters of simulated cash flows to reduce solutions' dependence on $X_{i,t}$ at time $t = 0$. There are 148 quarters (37 years of data) thus, $\tilde{T}_i$ is set to 148.

C Whole-sample estimation vs. Firm-level estimation

For a concrete illustration of these two estimation procedures, please consider a heterogeneous parameter case where there are two firms with true κ as: $\kappa_1 = 0.1$ and $\kappa_2 = 0.3$. Let

[14] I make sure that every firm face the same 50 time-series of aggregate earning growth.

us think of a no-omitted variable case ($\sigma_\epsilon = 0$). Then, observable leverages are $l_1 = f(0.1)$ and $l_2 = f(0.3)$. W estimation would get us $\hat{\kappa} = f^{-1}((l_1 + l_2)/2) = 0.1342$ and thus a bias due to heterogenenous parameter is $0.2 - 0.1342 = 0.0658$. As illustrated in Figure 3, W estimate is always downward-biased. Now, let us consider a heterogeneous parameter case where there are four firms with true κ as: $\kappa_1 = 0.1$, $\kappa = 0.15$, $\kappa_2 = 0.25$ and $\kappa = 0.3$. Then, W estimation gets us $\hat{\kappa} = f^{-1}((l_1 + l_2 + l_3 + l_4)/4) = 0.1527$ and thus the bias is 0.0473. This illustrates that bias exists in the small sample decreases in magnitude as the sample size increases. But F estimation gets us 0.2 for both scenarios and thus no bias. This illustrates that F estimation does significantly better than W when there is no omitted variable problem.

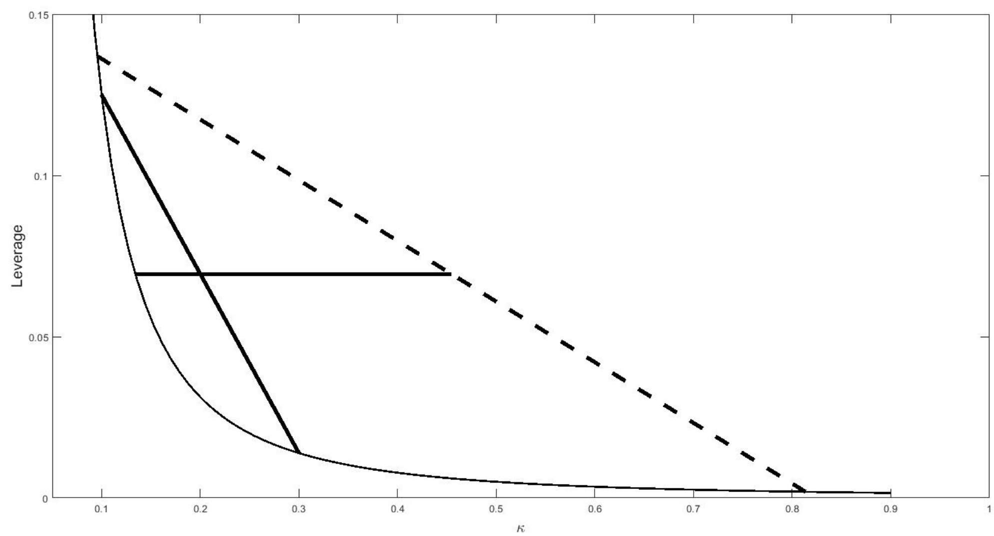

Figure 3: Whole-sample estimation vs. firm-level estimation

However, when there is omitted variable problem, W estimate suffers from smaller bias than F estimation does. Let us think of a scenario where observable leverages for two firms are $l_1 = f(0.1) + 0.012$ and $l_2 = f(0.3) - 0.012$. Then, W estimation gets us 0.1342 whereas F estimate gets us 0.4545.

This toy example tells us that W estimate does not suffer from omitted variable problem and heterogeneity problem as long as sample size is sufficiently large. But F does suffer from omitted variable and/or heterogeneity problem even when sample size is large.

D Estimation Results

	Data	Simulation
Book Lev	0.276	0.246
CAPM-β	1.136	1.693
PE	3.570	3.541
Earning Growth	0.005	0.002
Earning Growth Squared	0.033	0.036
Default Prob	0.040	0.039

Table 15: Data and Model-implied Moments for Restricted Model.

Table 16 documents sample average for main moments for each bucket.

	1	2	3
Data: book leverage			
Size	0.260	0.285	0.278
Tangibility	0.279	0.265	0.278
Intangibility	0.309	0.264	0.248
Data: default probability (%)			
Size	4.718	4.205	3.182
Tangibility	3.352	3.767	4.753
Intangibility	4.895	3.485	3.556

Table 16: Moments for each bucket