Empirical Corporate Finance in a Dynamic World

Arthur Korteweg Ilya Strebulaev*

May 6 2010

VERY PRELIMINARY AND INCOMPLETE
PLEASE DO NOT CITE

Abstract

The traditional regression methodology of estimating "optimal" capital structure is profoundly misspecified if firms follow dynamic financial policies. Instead, capital structure researchers, under many reasonable scenarios, should estimate the conditional *mode* of the leverage distribution. This metric is theoretically motivated and robust to a wide range of modeling assumptions. We introduce an empirical methodology to estimate the conditional mode of leverage and apply it to Compustat firms between 1950 and 2005. The estimated cross-sectional optimal leverage ratio decreases from a standard regression estimate of 37% to a mode estimate of 25%. Our findings should apply to most other areas of corporate finance research and call for a widespread rethinking of empirical tests.

*Both authors are at the Graduate School of Business, Stanford University. We thank Mike Harrison and Stanford and London School of Economics finance brown bag participants for helpful discussions and comments.

Obtaining estimates of "optimal" or "target" leverage ratios is at the heart of most traditional (i.e. non-structural) empirical capital structure research as well as perhaps the most important link between empirical analysis and capital structure theory. Starting from classical early studies of Taggart (1977), Marsh (1982), Titman and Wessels (1988)[1] to most recent contributions in the field (e.g. Baker and Wurgler (2002), Fama and French (2002), Korajczyk and Levy (2003), Flannery and Rangan (2006), Kayhan and Titman (2007), Lemmon, Roberts, and Zechner (2008), Byoun (2008), Huang and Ritter (2009), Chang and Dasgupta (2009)), the goal has been either to estimate the cross-sectional determinants of corporate financial decisions and compare the results with predictions of various theories, or to use such an estimate as the first step to conducting further investigation such as the analysis of mean-reversion in leverage. This empirical search for "target" leverage ratios is well-grounded in a number of static models of corporate financial structure, which contain many comparative statics predictions.

In this paper we show that traditional empirical methodology, in many realistic situations, yields highly misleading implications about even such a "simple" parameter as the average cross-sectional target leverage ratio. We claim that the traditional methods cannot be used to make any empirical statements about the extent of leverage or its determinants or as a qualitative link between empirics and theory. In short, we suggest the need to rethink the way traditional empirical methods are used and we propose an alternative way to study target capital structure that is more robust, yet not substantially more complicated. As will become obvious, our critique as well as our suggested remedy are not confined to capital structure but can be directly applied to other areas of financial economics.

To understand the crux of the matter, note that the standard econometric methodology in the extant literature is to estimate conditional expected leverage ratios, $E(L|X)$, where L is the leverage ratio and X is the set of determinants, using method-of-moment estimators, either OLS or GMM. These estimates are then used to pro-

[1]For a comprehensive overview of classical empirical studies in capital structure, see Harris and Raviv (1991).

vide insights about target capital structure. Unfortunately, in a dynamic world this method works only if firms readjust to the target instantaneously, an unlikely scenario. We show that in a dynamic world with a fixed transaction costs component (or, more generically, if firms do not re-adjust leverage to the target leverage ratio continuously), an estimate of $E(L|X)$ is *not* the target leverage ratio. Instead, under many reasonable scenarios, one should use the conditional mode of the leverage distribution, which we denote as $m(L|X)$.

This apparently minor distinction (from the economic point of view, especially for those versed in static models) turns out to be crucial because the dynamic behavior of individual firms produces cross-sectional distributions which are highly asymmetric around the target leverage ratio. To see the implications, consider a simple world where all firms have identical target leverage ratios. Even though their target is the same across firms, firms will have different leverage in the cross-section if they do not instantaneously mean-revert and there is an imperfect correlation in shocks across firms (both features are highly plausible, the pre-requisite for the latter being the presence of firm-specific idiosyncratic shocks). There are many reasons why in this case the cross-sectional distribution will be asymmetric around the target leverage ratio. One reason (perhaps not the most important but the easiest one to identify intuitively; more reasons will be identified later) is that firms have non-zero growth rates, so for example a firm with a positive growth rate is likely to have more frequently a lower leverage ratio (measured relative to its target) than a higher leverage ratio. It turns out that the mode of this distribution in most cases *is* the target leverage ratio and the mean *is not*. That the mean is not equal to the target leverage ratio should follow from the asymmetry. Therefore, any estimates based on the "conditional mean" are biased, and often have little or no economic meaning in a dynamic model.

How large is the bias? Of course, it depends on the values of too many real-life parameters to claim any unequivocal conclusions, and the results are subject to discussion and interpretation, but a surprising result of our work is that for *too many* sets of parameters the bias is uncomfortably large. For example, consider a firm whose earnings grow at an annual rate of 1% with 25% volatility, and which rebalances to an assumed optimal leverage ratio of 30% when leverage hits 20%, and which files for bankruptcy when leverage is 100% (i.e. equity is worthless at bankruptcy). As mentioned above, the mode of the cross-sectional distribution is the optimal leverage of 30%. However, the mean leverage ratio of this cross-sectional distribution is 52%, a difference of 22%!

To investigate whether such a difference between the mean and the mode is empirically plausible, we analyze all non-financial Compustat firms with positive leverage between 1950 and 2005. We find a mean leverage ratio of 37% compared to a mode of 25%. One consequence of this finding is that in order to come to grips with such a significantly lower target leverage ratio, either the costs of financial distress and/or net agency costs of debt must be larger than previously thought or there must be other economically substantial costs to leverage. We also find that asset tangibility is a more important determinant of capital structure than the traditional regression approach implies, whereas market-to-book ratios are significantly less important than usually thought. The importance of market-to-book in the traditional regression results has spurred a flurry of research based on dynamic debt overhang but out results suggest that theories explaining asset tangibility deserve more attention based on their relative empirical importance.

One result that empirical capital structure researchers working on non-structural estimations should take from our findings is that *it is incorrect to run cross-sectional or panel OLS regressions (or their more sophisticated moment-based cousins) with leverage ratio as the dependent variable.* This is a far-reaching result and begs the question what methods should replace the existing ones. We propose two related methods to estimate the conditional mode of the leverage distribution: i) a linear mode regression in the spirit of least squares, and ii) a non-parametric kernel method. Both methods are easy to use and yield quantitatively similar results. Their properties have been studied to some extent in the statistics and econometrics literature. In particular, the mode regression is a version of "robust regression" that has been applied in the economics and finance literature as a robustness test to OLS (e.g. Strömberg (2004) and Deschênes and Greenstone (2007)). The second method uses a kernel approximation to the conditional leverage density to estimate the mode. It is a non-parametric maximum likelihood estimator, which has been used in asset pricing applications (e.g. Stanton (1997)) but not as an alternative to OLS in corporate finance research. Because of its perceived importance for future research, we devote a separate section of this paper to introduce the major features of the econometric methodology using capital structure as an example so that empirical researchers can use it as a starting base for their analysis right away.

On a theoretical level, we show under what conditions the mode of the cross-sectional leverage distribution is the optimal leverage ratio upon refinancing. The natural benchmark case is based on a leverage process that is consistent with contingent-claims models of the firm in the Black-Scholes (1973) and Merton (1974) tradition, even though our results should go through for more general firm-value and cash-flow processes. The degree to which the average leverage ratio is above or below the optimum is determined by the growth rate of assets or cash flows and the relative propensity to refinance when the firm is above or below its target leverage ratio (i.e., its mean-reversion behavior).

The importance of fixed costs for economic applications has been emphasized in economics for many decades. In a similar fashion, it has been argued in dynamic capital structure models (e.g. Fischer, Heinkel and Zechner (1989), Goldstein, Ju and Leland (2001)) that the presence of fixed costs can generate a large spread in leverage ratios. Strebulaev (2007) showed that the non-trivial dynamic cross-sectional behavior of leverage indeed can be generated by relatively small frictions and Korteweg (2010) gives empirical evidence of the size of friction costs. We assume, following the multitude of theoretical literature, that firms follow some optimal capital structure. The economic forces behind determining the optimal leverage ratio are not particularly important to us as long as they lead to specifying a certain target leverage ratio that firms would like to follow. It is important to stress that while in many existing contingent-claims models the optimal capital structure is the result of the trade-off between the present value of debt tax shields and bankruptcy costs, our results do not rely on the exact economic mechanism (other reasons may include an attempt to resolve various agency problems, information asymmetry, product market competition, and so on and so forth), as long as that mechanism leads to a target leverage ratio.

We should note that at this stage we do not consider incomplete rebalancing i.e. firms refinancing towards the target only partially. In this setting leverage ratios even at restructuring may not equal long-term target leverage ratios, and we consider this an important issue for future research.

We wish to stress that, although we present our results only for capital structure, its implications are likely to be applicable to a much broader range of topics as long as the dynamic behavior of agents (firms, consumers, investors, governments) leads to an asymmetric cross-sectional distribution of the variables of interest. Examples of such topics are cash flow and payout policies, corporate investment policy with fixed adjustment costs, or government economic policy under which intervention occurs only when a certain indicator such as unemployment rises above a threshold. Average investment or average unemployment are meaningless in these settings and alternative measures akin to the ones in this paper need to be developed.

The paper is structured as follows. Section I describes our model of the firm and characterizes the leverage distribution. Section II introduces our empirical methodology. Section III describes the data, and Section IV discusses the empirical results. Section V shows robustness and Section VI concludes.

I Model

A A simple model

To establish the intuition of our main results, consider the following toy model of the firm leverage. Although the model is very simple, most of the results borne out by the model will hold in our benchmark model as well. Assume that firms follow target leverage policy but because of imperfections, such as fixed transaction costs, they deviate from their target leverage ratio.[2] Define the firm's target leverage ratio as L^* and assume that it is a constant.[3] Intuitively, the firms follow a so-called (s, S) policy – they are inactive if the leverage, due to market fluctuations, deviates a little and they refinance when the shocks are too large, i.e. when leverage deviates too much from L^*. Define a lower refinancing boundary by L^d and an upper refinancing boundary by L^u, $0 \leq L^d \leq L^* \leq L^u \leq 1$. As long as $L_t \in (L^d, L^u)$, the firm is in the inaction region: it does not take any actions to change its leverage. The moment that leverage reaches one of the boundaries, the firm refinances, as illustrated in Figure 1. We consider the case of "complete" refinancing, by which we mean that the firm refinances to the optimal leverage ratio L^* whenever it hits either boundary, a case considered in most dynamic capital structure models to date, such as Fischer, Heinkel and Zechner (1989), Goldstein, Ju, and Leland (2001), and Dangl and Zechner (2003).[4]

Assume that the firm's leverage ratio follows a Geometric Brownian Motion (GBM) in the inaction regime, i.e., *in between refinancings*:

$$dL_t = \mu L_t dt + \sigma L_t dW_t, \tag{1}$$

where the constant μ is the instantaneous growth rate of leverage, the constant σ is the instantaneous volatility, and dW is the Wiener process under the actual measure. Even though this may look as an ad hoc haphazard model of leverage, in fact dynamic contingent claims capital structure models lead to the same structure, with the major difference that parameters μ and σ are state-dependent.

[2]Note that for our purposes the optimality of "target" leverage as well as other parameters of interest is immaterial.

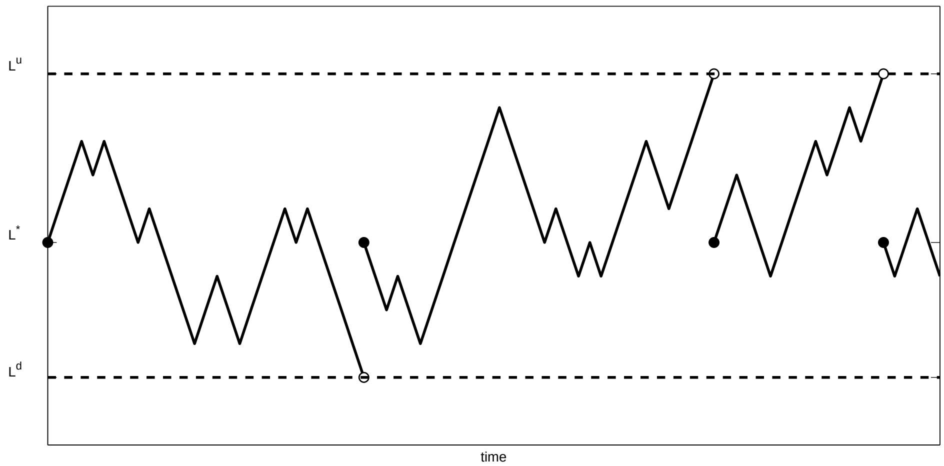
[3]Theoretically, the market leverage ratio (the ratio of the market value of debt to the sum of the market values of debt and equity) is used. Empirically, market values of debt are not always available and therefore the so-called quasi-market leverage ratio (the ratio of the book value of debt to the sum of the book value of debt and the market value of equity) is used. For our purposes, the distinction is not important and in this section we do not specify "leverage ratio" any further.

[4]In many models, the upper refinancing boundary implies that the firm enters financial distress and defaults. This is a partial case of our framework, when $L^u = 1$, assuming that the firm levers up to the same target ratio once it leaves the bankruptcy process.

Figure 1

Example of time-series of leverage

This figure illustrates the rebalancing of leverage at the boundaries L^d and L^u.



Proposition 1 derives the stationary cross-sectional distribution of leverage under these assumptions.

Proposition 1 Stationary distribution of leverage.

The stationary distribution of leverage when the leverage follows the process (1) in the inaction region $L \in (L^d, L^u)$ and the leverage re-starts at L^* at both triggers L^d and L^u is:[5]

$$f(L) = \left[\frac{L^d}{L} - \left(\frac{L^d}{L}\right)^{-k+2}\right] / C_1 \quad for \quad L^d < L \leq L^* \tag{2}$$

$$f(L) = \left[\frac{L^u}{L} - \left(\frac{L^u}{L}\right)^{-k+2}\right] / C_2 \quad for \quad L^* < L < L^u, \tag{3}$$

[5]The proposition omits a special and uninteresting case, where $\mu = \frac{1}{2}\sigma^2$. The details are available from the authors upon request.

for $k \neq 1$, where $k \equiv 2\mu/\sigma^2$ and

$$C_1 \equiv L^d \left[-\log\left(\frac{L^d}{L^*}\right) + \log\left(\frac{L^u}{L^*}\right) \cdot \frac{1-(\frac{L^d}{L^*})^{-k+1}}{1-(\frac{L^u}{L^*})^{-k+1}} \right] \quad (4)$$

$$C_2 \equiv L^u \left[-\log\left(\frac{L^d}{L^*}\right) \cdot \frac{1-(\frac{L^u}{L^*})^{-k+1}}{1-(\frac{L^d}{L^*})^{-k+1}} + \log\left(\frac{L^u}{L^*}\right) \right]. \quad (5)$$

Figure 2 shows a typical example of this distribution, assuming $L^d = 0.1$, $L^u = 0.6$, $L^* = 0.2$, $\mu = 0.05$ and $\sigma = 0.4$. Several features stand out. First, the distribution is asymmetric, skewed to the right, and the density is non-differentiable at the mode. Second, the mode of the distribution equals the optimal leverage ratio, L^*. Third, in this particular exercise, the mean of the distribution is 0.24, higher than the target leverage ratio.

The comparative statics of the cross-sectional distribution is intuitive. Higher values of the growth rate μ make higher leverage more likely and shift the distribution to the right of L^*. Higher values of the volatility σ tone down the drift effect as it increases the probability of leverage moving in the opposite direction of the drift. Higher values of the lower refinancing boundary L^d and the upper refinancing boundary L^u imply that the firm spends more time in the higher-leverage region, shifting the distribution to the right of L^*.

The first important result, illustrated in Figure 2, is that the target leverage is the *mode* of the distribution, which we denote as $m(L)$. How general is this condition? Proposition 2 finds the mode of the distribution $f(L)$:

Proposition 2 THE MODE OF THE DISTRIBUTION. *The mode of the distribution $f(L)$ is:*

$$m(L) = \begin{cases} L^*, & \text{if } k \geq 2, \\ \min\{L^*, \left(\frac{1}{2-k}\right)^{\frac{1}{k-1}} L^d\}, & \text{if } k < 2. \end{cases} \quad (6)$$

It follows that the target leverage ratio L^* is the mode of the cross-sectional leverage distribution either if $\mu > \sigma^2$ or if $L^* < \left(\frac{1}{2-k}\right)^{\frac{1}{k-1}} L^d$. The latter condition is satisfied whenever μ or L^d are sufficiently large. For example, for the value of parameters in

Figure 2

Stationary leverage distribution: Geometric Brownian Motion

This figure shows the stationary distribution of leverage, with lower and upper refinancing boundaries $L^d = 0.1$, $L^u = 0.6$, and optimal leverage $L^* = 0.2$. The leverage process follows a Geometric Brownian Motion in between refinancings, with $\mu = 0.05$ and $\sigma = 0.4$.

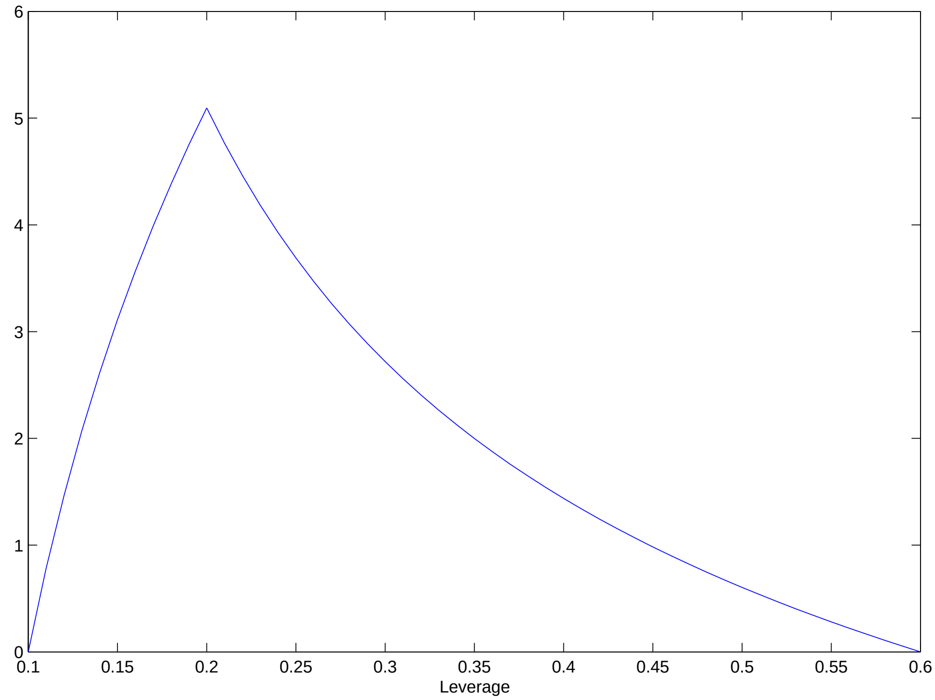


Figure 2 L^* is the mode as long as $\mu \geq 0$ or $L^d \geq 0.0856$. In short, this condition is likely to hold for most reasonable scenarios.[6] [7] Intuitively, when μ is very small or when L^d is very small, firms are likely to spend more time in the low-leverage region and rarely return to the target leverage ratio so that the mode can theoretically be lower than the target leverage ratio. However, the proposition implies that the mode

[6]Interestingly, if the leverage follows a standard (not geometric) Brownian Motion, the mode is *always* equal to L^*. The details are available upon request.

[7]With E expected to increase one might suspect that L is expected to drift downwards. However, in many such scenarios leverage in fact drifts upwards due to Jensen's Inequality, since L is convex in E.

can never be higher than the target leverage ratio. This gives an important upper bound on the cross-sectional target leverage ratio.

The intuition behind the mode result is that the optimal leverage ratio is visited most frequently because not only does the firm pass it by as any other leverage ratio in between refinancings, it is also the ratio that is adopted whenever the firm refinances. Therefore, the leverage ratio that the firm shows most often is exactly the target leverage ratio.

The second important result, illustrated in Figure 2, is that the cross-sectional average leverage ratio, which is equal to 0.24, is higher than the target leverage ratio and the mode. This is inconsistent with the traditional applications of the standard dynamic capital structure models, which was noted by Strebulaev (2007). Because the cross-sectional distribution is generically asymmetric, we should expect the mean leverage ratio to be generally different from L^*, barring knife-edge type solutions. Proposition 2 implies therefore that in many, if not most, realistic scenarios, the average leverage ratio *cannot* be the target leverage ratio. This is an important result because it shows that the empirical estimates of the average leverage ratio in the cross-section are very unlikely to be informative about the optimal financial policies firms. The mean of the cross-sectional distribution can be above or below L^* depending on the parameter choice. To study the relation of these to quantities better, Proposition 3 gives the value of the mean:

Proposition 3 THE MEAN OF THE DISTRIBUTION. *The mean of the distribution $f(L)$ is:*

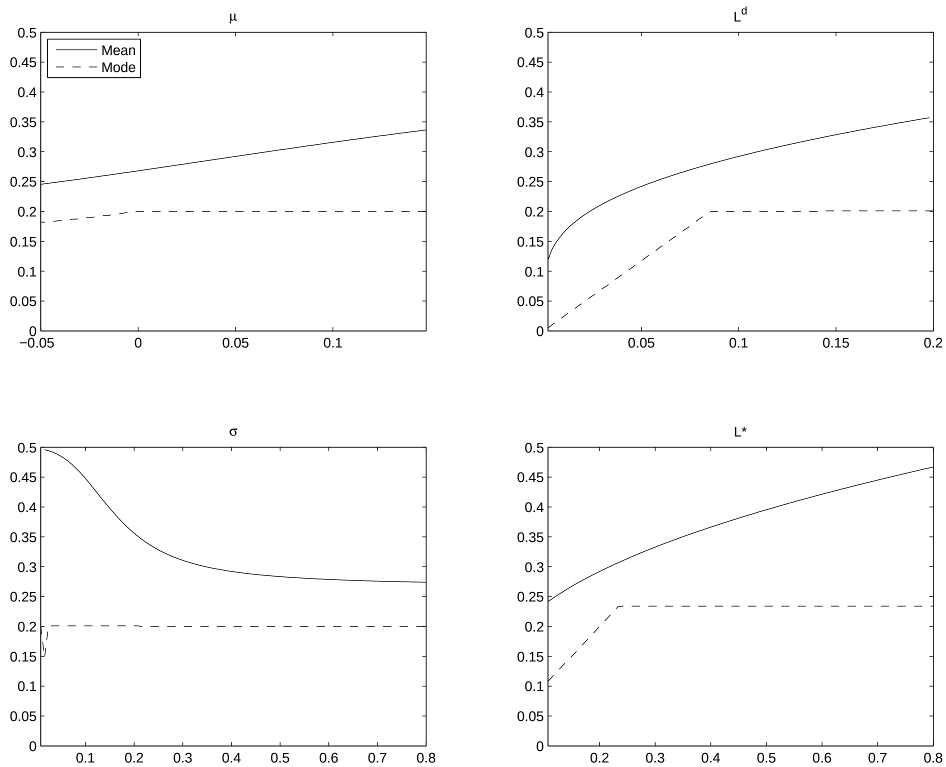
$$\begin{aligned} E(L) &= \frac{L^d}{C_1}\left(L^* - L^d\right) + \frac{L^u}{C_2}\left(L^u - L^*\right) + \\ &\quad \frac{(L^d)^2}{kC_1}\left[1 - \left(\frac{L^d}{L^*}\right)^{-k}\right] - \frac{(L^u)^2}{kC_2}\left[1 - \left(\frac{L^u}{L^*}\right)^{-k}\right]. \end{aligned} \tag{7}$$

To explore the relation between the mean and the mode, Figure 3 shows the values of these two quantities as the value of the primitive parameters changes. In the benchmark case, we set $\mu = 0.05$, $\sigma = 0.4$, $L^d = 0.1$, $L^u = 1$, and $L^* = 0.2$.

Figure 3

Comparative Statics: Geometric Brownian Motion

This figure shows the mean and mode of the stationary distribution of leverage. For each plot, one parameter is allowed to vary (on the horizontal axis), and the remaining parameters are kept at default values: $\mu = 0.05$, $\sigma = 0.4$, $L^d = 0.1$, $L^* = 0.2$ and $L^u = 1$.



Each plot in the figure shows comparative statics as a single parameter value varies while the remaining parameters are kept at default values. The results illustrated by the figure can be summarized as follows: the average cross-sectional leverage ratio is always larger than the mode and in many cases is also larger than the optimal leverage ratio, in the benchmark case by about 9%. The bias becomes larger for larger values of the growth rate, the lower leverage trigger, and for smaller values of the volatility parameter. For the benchmark set of parameters, the mode is the optimal leverage ratio. The mode is below the optimal leverage ratio when either the growth rate is very low or when the difference between L^d and L^* is too large. As we show in the next section, in the context of the contingent-claim model of capital structure, which results in a more realistic stochastic leverage process, the comparative statics results show the mode in a much more favorable light even though the main intuition is preserved.

B A benchmark dynamic model of capital structure

In this section we re-consider the results derived in Section A for a benchmark model in the standard contingent claims framework of optimal dynamic capital structure based on the trade-off considerations. It should be intuitively clear that our results are qualitatively robust to various assumptions and extensions, as well as to the precise nature of the economic mechanism leading to the optimal target leverage, but to get quantitative predictions we consider the simplest model in the framework which admits tractable solutions. The model follows the previous work such as of Fisher, Heinkel, Zechner (1989) and Goldstein, Ju, and Leland (2001).

To emphasize, our model does not impose a single assumption that is not present in the extant models, even though we simplify some assumptions for clarity. The main economic assumption is that adjustment costs lead the firm to follow an impulse control policy by adjusting its leverage back to the optimal refinancing ratio in a lumpy way only when its cash flow reaches lower or upper trigger points. The main technical assumption is that the variables are chosen in such a way that the first-order homogeneity (i.e. the ubiquitous scaling property) is preserved which enables us to solve an otherwise high-dimensional dynamic problem.

A single firm's cash flow, δ_t, follows, under the actual probability measure, a stochastic process:

$$d\delta_t = \mu_t^P dt + \sigma_t dW_t^P, \tag{8}$$

where μ_t^P is the instantaneous growth rate under the actual measure, σ_t is the instantaneous volatility of the growth rate, and dW_t^P is a Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$.

To follow the standard assumptions, we specify further by assuming that $\mu_t^P = \mu^P \delta_t$ and $\sigma_t = \sigma\delta_t$, where μ and σ are constants. Therefore, the cash flow process follows the geometric Brownian motion

$$d\delta_t = \mu^P \delta_t dt + \sigma \delta_t dW_t. \tag{9}$$

This process is usually called the EBIT-process, emphasizing that the value of δ represents is the pre-tax earnings. For our purposes, this is of less importance.

For the pricing purposes, we will also require the cash flow process under the risk-neutral measure $\mathbb{Q}$, which follows

$$d\delta_t = \mu \delta_t dt + \sigma \delta_t dW_t^Q, \tag{10}$$

where μ is the instantaneous growth rate under the risk-neutral measure and dW_t^Q is a Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{Q}, (\mathcal{F}_t)_{t\geq 0})$.

Equityholders maximize their value by following optimal financial policy in the presence of the trade-off between tax benefits from debt and costs of default. Two securities are allowed: common stock and debt, which takes the form of a callable perpetuity (an assumption of noncallable debt that may dilute yields similar results) with coupon level c, chosen by the firm.

Assume that the proportional tax rate on corporate earnings is τ, the proportional debt issuance cost is γ, the proportional bankruptcy cost is α, and the riskless interest rate is r.

B.1 Contingent claims values and leverage

In the presence of adjustment costs, the firm will adjust its leverage infrequently. Let δ_0 be the initial cash flow level, and, without any loss of generality, we may assume that $\delta_0 = 1$. Then the firm will refinance whenever the cash flow level either reaches the upper threshold, δ_R, $\delta_R > 1$, or a lower threshold, δ_B, $\delta_B < 1$. The notation reflects the intuition that δ_R is the refinancing point and δ_B could be the point of bankruptcy, default, or distress, as in Goldstein, Ju, Leland (2001), Strebulaev (2007), and other work, although this should not necessarily be the case. The model will provide specific values for optimal δ_R and δ_B, but for our results the only fact that will be important about these optimal parameters is that both of them could be reached in finite time, so that $\delta_R < \infty$ and $\delta_B > 0$.

To summarize, the firm has three controls at its disposal: the coupon level c, the upper trigger cash flow level δ_R and the lower trigger cash flow level δ_B. At either trigger point, the firm refinances (at δ_B we can assume that debtholders take over and refinance the new firm) and, because of the scalability property (achieved by the combination of the geometric Brownian motion, perpetuity form of debt, and the nature of debt issuance costs), it is a proportionally smaller or larger copy of itself at the previous refinancing.

To derive the market values of equity and debt, we start by introducing the primitive Arrow-Debreu-type securities in this economy. Specifically, define $A(\delta, \delta_B, \delta_R)$ to be the value of the security that pays \$1 when the bankruptcy trigger δ_B is reached before the refinancing trigger δ_R conditional on the current value of δ (we will assume henceforth that $\delta_B < \delta < \delta_R$). Analogously, define $A(\delta, \delta_R, \delta_B)$ as the value of the security that pays \$1 when δ_R is reached before δ_B.

It is well known that the value of these Arrow-Debreu securities can be written as follows:

$$A(\delta, \delta_B, \delta_R) = \frac{\delta^x \delta_R^y - \delta_R^x \delta^y}{\delta_B^x \delta_R^y - \delta_R^x \delta_B^y}, \tag{11}$$

$$A(\delta, \delta_R, \delta_B) = \frac{\delta_B^x \delta^y - \delta^x \delta_B^y}{\delta_B^x \delta_R^y - \delta_R^x \delta_B^y}, \tag{12}$$

where x and y are:

$$x = -\frac{1}{\sigma^2}\left(\mu - \frac{\sigma^2}{2} - \sqrt{\left(\mu - \frac{\sigma^2}{2}\right)^2 + 2\sigma^2 r}\right), \tag{13}$$

$$y = -\frac{1}{\sigma^2}\left(\mu - \frac{\sigma^2}{2} + \sqrt{\left(\mu - \frac{\sigma^2}{2}\right)^2 + 2\sigma^2 r}\right). \tag{14}$$

All other contingent claims in this economy could be expressed using the value of these securities. The market value of equity can be written as:

$$
\begin{aligned}
E &= (1-\tau)\left(\frac{\delta}{r-\mu}-\frac{c}{r}+A(\delta,\delta_B,\delta_R)\left(-\frac{\delta_B}{r-\mu}+\frac{c}{r}\right)\right) \\
&\quad +A(\delta,\delta_R,\delta_B)\left[(1-\tau)\left(-\frac{\delta_R}{r-\mu}+\frac{c}{r}\right)-B+\delta_R(1-\gamma)B+\delta_R E_0,\right] \qquad (15)
\end{aligned}
$$

where E_0 is the value of equity at the moment of refinancing (but after new debt is issued), i.e. when $\delta_0 = 1$. This value is found by equating E and E_0. Also, B is the book value of debt. It equals the market value of debt at the moment of refinancing, i.e. $B = B(\delta_0, c)$ and it does not depend on the current value of δ.

To understand the expression above, note that the first term, $\frac{\delta}{r-\mu} - \frac{c}{r}$, is the pre-tax equity value in the world without refinancing and bankruptcy. The second term is the adjustment for bankruptcy, upon which the equity foregoes the right to future cash flows and does not have to pay coupons any longer. The final term is the adjustment for refinancing, which can be thought of in the following way. The equity exchanges all its current rights and obligations for the future equity value. The first term in the square brackets represents the after-tax cash flows from the old equity which shareholders forego. In addition, the equity has to recall debt at its original value, B. In replacement, shareholders issue new debt at the cost γ. Because of the scalability, the value of the new debt issued is $\delta_R B$. After the debt is issued, equityholders also get the residual value of equity, $\delta_R E_0$.

Define G_E as:

$$
\begin{aligned}
G_E(\delta) &= (1-\tau)\left(\frac{\delta}{r-\mu}-\frac{c}{r}+A(\delta,\delta_B,\delta_R)\left(-\frac{\delta_B}{r-\mu}+\frac{c}{r}\right)\right) \\
&\quad +A(\delta,\delta_R,\delta_B)(1-\tau)\left(-\frac{\delta_R}{r-\mu}+\frac{c}{r}\right). \qquad (16)
\end{aligned}
$$

Then the value of equity at refinancing is equal to:

$$
E_0 = \frac{G_E(1) - A(1,\delta_R,\delta_B)B\left(1-\delta_R(1-\gamma)\right)}{1-A(1,\delta_R,\delta_B)\delta_R}. \qquad (17)
$$

Because B is the par value of debt, when $\delta_0 = 1$, the book value of B can similarly be expressed as:

$$B = \frac{c}{r} + A(1, \delta_B, \delta_R) \left(-\frac{c}{r} + (1-\alpha) \left(\delta_B((1-\gamma)B + E_0) \right) \right) + A(1, \delta_R, \delta_B) \left(-\frac{c}{r} + B \right). \tag{18}$$

The first term is the value of riskless perpetuity. The second term is the adjustment for bankruptcy. The debtholders forego the right to coupon payments but instead become the new equityholders. Note that we assume that new equityholders optimally lever up at default and choose the same optimal leverage ratio. Therefore, the value of the new firm is $\delta_B((1-\gamma)B + E_0)$, adjusted for bankruptcy costs. Alternatively, if they do not lever up, the second term would change to $A(\delta, \delta_B, \delta_R) \left(-\frac{c}{r} + (1-\tau)(1-\alpha)\frac{\delta_B}{r-\mu} \right)$. Although the former assumption is easier in our case because we can keep the number of the firms in the cross-sectional distribution constant, the results on optimal financial policies are very little affected by the assumption on what happens at default. Finally, the third term is the adjustment for refinancing when the debtholders forego future coupon payments and get their principal back.

From equations (17) and (18), the value of B can be written as:

$$B = \frac{\frac{c}{r}\left(1 - A(1, \delta_B, \delta_R) - A(1, \delta_R, \delta_B)\right) + \frac{A(1,\delta_B,\delta_R) G_E(1)}{1 - A(1,\delta_R,\delta_B)\delta_R}}{1 - A(1, \delta_B, \delta_R)(1-\alpha)\delta_B(1-\gamma) + \frac{A(1,\delta_R,\delta_B)A(1,\delta_B,\delta_R)(1-\delta_R(1-\gamma))}{1-A(1,\delta_R,\delta_B)\delta_R} - A(1, \delta_R, \delta_B)}. \tag{19}$$

The quasi-market leverage, $L(\delta)$, is defined at any point in time within the initial refinancing cycle as:

$$L(\delta) = \frac{B}{E(\delta) + B}. \tag{20}$$

Note that because we are interested entirely in the leverage distribution properties, the homogeneity feature allows us to rescale the level at any next refinancing to δ_0, so that this way of defining leverage applies equally to any point in time. We will study the behavior of the quasi-market leverage because empirically the market value of leverage is usually not available, but the results are, in fact, remain almost unchanged if we consider instead the market value of leverage.

We assume that the conflict of interests prevents equityholders from internalizing outstanding debtholders's value in choosing a default boundary, and therefore δ_B is found from maximizing the value of equity alone. This leads to a smooth-pasting condition

$$0 = \frac{\partial E}{\partial \delta}|_{\delta=\delta_B} = \frac{\partial G_E(\delta)}{\partial \delta} + \frac{\partial A(\delta, \delta_R, \delta_B)}{\partial \delta} \left[-B + \delta_R(1-\gamma)B + \delta_R E_0\right] \tag{21}$$

for any feasible values of c and δ_R. Finally, the optimization problem is for equityholders to maximize the value of the firm at date 0, at $\delta_0 = 1$:

$$\max_{c,\delta_R} E_0 + (1-\gamma)B \tag{22}$$

subject to the optimal choice of δ_B.[8]

B.2 Analysis of the model

We derive the stationary distribution of δ analogously to the stationary distribution of leverage when leverage follows a GBM. Assuming that when a firm refinances we rescale δ back to δ_0 and when it defaults a new firm enters at δ_0:

$$f(\delta) = \left[\frac{\delta_B}{\delta} - \left(\frac{\delta_B}{\delta}\right)^{-k+2}\right]/C_1 \quad \text{for} \quad \delta_B < \delta \leq \delta_0 \tag{23}$$

$$f(\delta) = \left[\frac{\delta_R}{\delta} - \left(\frac{\delta_R}{\delta}\right)^{-k+2}\right]/C_2 \quad \text{for} \quad \delta_0 < \delta < \delta_R, \tag{24}$$

for $k \neq 1$, where $k \equiv 2\mu/\sigma^2$ and

$$C_1 \equiv \delta_B \left[-\log\left(\frac{\delta_B}{\delta_0}\right) + \log\left(\frac{\delta_R}{\delta_0}\right) \cdot \frac{1-(\frac{\delta_B}{\delta_0})^{-k+1}}{1-(\frac{\delta_R}{\delta_0})^{-k+1}}\right] \tag{25}$$

$$C_2 \equiv \delta_R \left[-\log\left(\frac{\delta_B}{\delta_0}\right) \cdot \frac{1-(\frac{\delta_R}{\delta_0})^{-k+1}}{1-(\frac{\delta_B}{\delta_0})^{-k+1}} + \log\left(\frac{\delta_R}{\delta_0}\right)\right]. \tag{26}$$

Since leverage is a monotonically decreasing function of δ we can derive the stationary distribution of L:

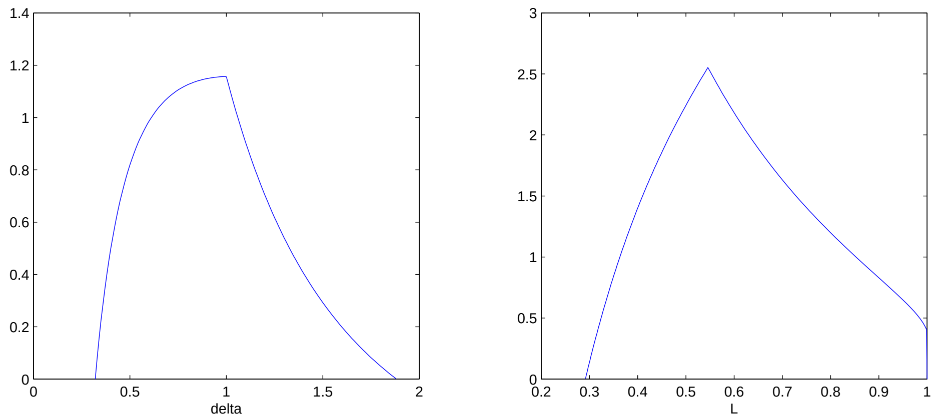
$$f_L(L) = -f_\delta(\delta(L)) \cdot \frac{\partial}{\partial L}\delta(L)$$

[8]We assume, following Goldstein, Ju, and Leland (2001), that equityholders internalize debtholders' value in deciding on the refinancing boundary. Although the conflict of interest in choosing a refinancing boundary may lead to interesting implications for optimal capital structure, this assumption is of no importance to our study.

Figure 4

Stationary leverage and δ distribution: The benchmark model

This figure shows the stationary distribution of leverage and cash flows for the benchmark model and for the following parameter values: $\mu = 0.01$, $\sigma = 0.25$, $\tau = 0.2$, $\alpha = 0.1$, $\gamma = 0.01$, $r = 0.05$, and $\mu^P = 0.06$.



where $\delta(L)$ is the inverse function of $L = B/(B + E(\delta))$.

Figure 4 shows a stationary distribution of δ and leverage for one set of parameters. The properties of the leverage distribution are similar to the distribution, when the leverage follows the GBM. Because the stochastic process of leverage is further away from the GBM as the leverage drifts further from the optimal leverage, the difference becomes more visible in the tails of the distribution (for example, a non-trivial fraction of firms have a very high leverage, though in the simple model the density attenuates to zero). In this case, the optimal leverage ratio, of 55%, equals the mode, and the mean, of 65%, is substantially higher.

To understand the relation between the optimal leverage, the mode, and the mean, we need to understand better the properties of the leverage stationary distribution. In particular, if $f'_L > 0$ for $L < L^*$ and $f'_L > 0$ for $L > L^*$, then the mode always equals the optimal leverage ratio. The derivative of f_L is:

$$f'_L(L) = -f'_\delta(\delta(L)) \cdot \left[\frac{\partial}{\partial L}\delta(L)\right]^2 - f_\delta(\delta(L)) \cdot \frac{\partial^2}{\partial L^2}\delta(L)$$

For $L > L^*$ the first term (including the negative sign) is negative (since $L > L^*$ corresponds to $\delta < \delta^*$). For $L < L^*$ the first term (including the negative sign) is positive. However, the conditions for the convexity of $\delta(L)$ (which determines the second derivative of $\delta(L)$ in the expression) is more challenging to establish. We have checked that $\delta(L)$ is convex over most of the space (except when L^* is very close to 1) for many parameter values, and for now present the comparative statics around the benchmark set of parameters that are used in Figure 4.

Figure 5 shows the comparative statics of the mean and mode of the leverage distribution in the model. As can be seen, the mode is generally very close to the optimal leverage at refinancing, while the mean is generally non-trivially larger. The figure shows that the mode performs much better in the contingent claims model than in the simple model, when leverage is unrealistically set to follow the GBM. There are several reasons for that. First, the contingent claims model is more parsimonious, because it allows us to disentangle between risk-neutral and actual measures and introduce a risk premium. Second, the simple model performed particularly badly when the difference between the optimal leverage ratio and the lower leverage trigger were large. In the benchmark model, however, the difference is a function of the underlying parameters and the optimal leverage ratio and the lower leverage trigger typically move in the same direction, preserving the magnitude of the difference.

Our preliminary evidence therefore strongly suggests that the mode is likely to be equal to the optimal leverage ratio at refinancing in the stationary cross-section for most realistic scenarios, and that the mean leverage ratio is not. Moreover, the relation between the mean, on the one hand, and the mode and L^*, on the other hand, changes substantially as the fundamental parameters vary.

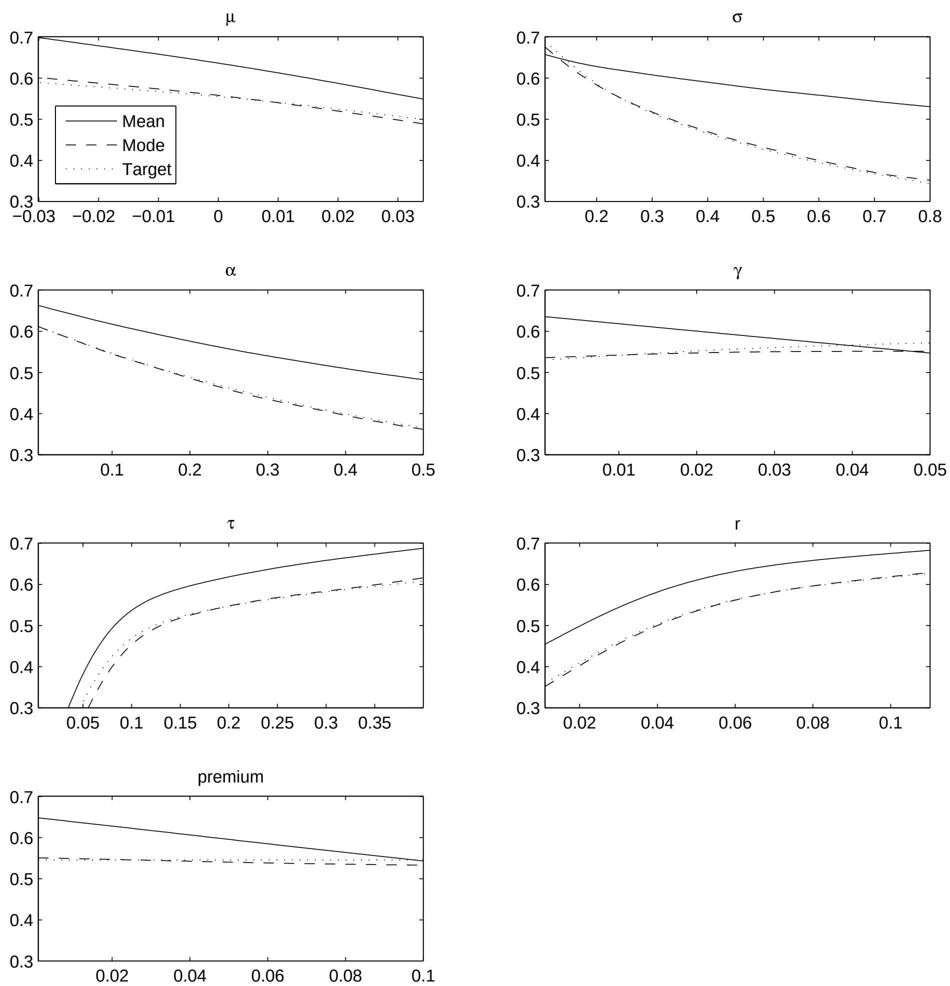
B.3 Discussion

There are several important caveats to these results: (1) This is the complete refinancing case; (2) We assumed that the firm is in a stationary setting, with fixed refinancing policies; (3) We assumed that there are no systematic shocks, i.e. all cash flow processes for the cross-section of firms are independent. Relaxing these assump-

Figure 5

Comparative Statics: Structural Model

This figure shows the mean and mode of the stationary distribution of leverage of the structural model. We also show the optimal leverage ratio. For each plot, one parameter is allowed to vary (on the horizontal axis), and the remaining parameters are kept at default values: $\mu = 1\%$, $\sigma = 25\%$, $\tau = 20\%$, $\alpha = 10\%$, $\gamma = 1\%$, $r = 5\%$, and $\mu_P = 0.06$.



tions and testing the robustness of our measure is an important avenue for future research.

II Mode Estimation

The previous section establishes the result that the (conditional) mode of leverage is generally the optimal capital structure upon refinancing, in contrast to the common practice of using the conditional mean estimated by least-squares regression or GMM. In this section we develop two empirical methods for estimating the conditional mode. The first method, mode regression, is akin to OLS but uses a different objective function. The second method is a non-parametric maximum likelihood approach (NPMLE) using a kernel approximation to the conditional distribution of leverage. We also discuss how the two methods are related.

A Mode Regression

The mode regression developed by Lee (1989, 1993) is based on work going back as far as Parzen (1962) and Chernoff (1964). The intuition behind mode regression is similar to OLS. Assume we have N draws $(L,x)_{i=1\ldots N}$ from the joint distribution of leverage, L, and a vector of J covariates, x. In OLS we assume that the conditional mean of leverage is linear in x, $E(L|x)=x'\beta$, and find β by minimizing the sum of squared errors. In a mode regression we assume that the conditional *mode* is linear in x, $m(L|x)=x'\beta$, and estimate β by minimizing the loss function:

$$Q=\sum_{i=1}^{N}1\left\{\left|\frac{L_i-x_i'\beta}{h}\right|\geq 1\right\}, \tag{27}$$

where $1\{\cdot\}$ is the indicator function, and h is the bandwidth.[9]

The loss function (27) is minimized at the middle value of the interval of length $2h$ around $x'\beta$ that contains the highest number of observations. In the limit this is the interval that captures the most probability under $f(L|x)$.

[9]In essence the mode regression is a specific implementation of Huber's (1972) M-class estimator, which maximizes $\sum_{i=1}^{N}\rho(L_i-x_i'\beta)$ for some function $\rho(\cdot)$. Here, $\rho=1\left\{\left|\frac{L_i-x_i'\beta}{h}\right|<1\right\}$.

Table I

Mode Regression with Sample Data

Panel A: Objective function Q for $\beta = 0.4$

L data	0.1	0.2	0.3	0.3	0.3	0.4	0.6	0.7	0.8	0.9
$L - x'\beta$	-0.3	-0.2	-0.1	-0.1	-0.1	0	0.2	0.3	0.4	0.5
$\left\|\frac{L_i - x_i'\beta}{h}\right\|$	6	4	2	2	2	0	4	6	8	10
$\left\|\frac{L_i - x_i'\beta}{h}\right\| \geq 1$	1	1	1	1	1	0	1	1	1	1

Panel B: Q as a function of β and h

	β									
h	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.05	9	9	7	9	10	9	9	9	9	10
0.1	9	6	6	9	8	8	8	8	8	9
0.2	5	5	5	5	7	7	7	7	8	8

A simple example illustrates the method. Suppose we observe the ten leverage ratios shown in the top line of Table I Panel A. We want to estimate the mode of this distribution. For now assume there are no covariates but only an intercept, $x = 1$, so that β is the mode. Suppose we guess that $\beta = 0.4$. The second line in Table I shows the deviation of observed leverage from the fitted mode. In the third line we calculate the absolute deviation, scaled by a bandwidth h of 0.05 (we will discuss the bandwidth choice in more detail below). The last line indicates whether an observation is inside ("0") or outside ("1") the bandwidth. For this choice of β and h nine observations are outside the bandwidth, hence $Q = 9$. Panel B of Table I shows the value of Q for a range of β and h. For $h = 0.05$ it is clear that Q is minimized at $\beta = 0.3$, which is indeed the mode in this dataset.[10] An alternative to the grid search to find the minimum of Q is to use standard optimizers that are built into statistical packages.

Dealing with covariates in the data is straightforward. The only twist to the above procedure is that x_i is now a vector, and we need to search over multiple betas to find

[10] Note that $\beta = 0.29$ or $\beta = 0.31$ would give the same minimized value of Q. With more datapoints, this problem disappears.

the minimum of Q. For example with a constant and one covariate in x, we have to find a set of two betas (the intercept and the loading on the covariate) to minimize Q. As above, this can be done on a grid (of two dimensions) or using standard optimizers.

The bandwidth $h > 0$ is chosen by the econometrician. A higher bandwidth increases the number of observations that are included in calculating Q and reduces the standard error. On the flipside, a higher bandwidth raises the bias in estimated β if $f(L|x)$ is asymmetric. This trade-off between precision and bias guides the choice of bandwidth. If we let h decrease with sample size, the estimator $\widehat{\beta}$ converges in probability to the true β that determines the conditional mode. Lee (1989) suggests using a bandwidth such that the proportion of datapoints that falls within $2h$ of the mode is around 70% (i.e. $1 - Q/N \approx 0.7$). For the empirical leverage data of section III, a good choice of bandwidth is 0.2.[11]

We illustrate the trade-off embodied in the choice of bandwidth with our numerical example in panel B of Table I. If we increase the bandwidth to 0.1 we include more observations (i.e. Q is lower for most choices of β), but we can no longer distinguish $\beta = 0.2$ from 0.3. Raising the bandwidth to 0.2 gives the same value of Q for β in the range 0.1 to 0.4. Choosing $\beta = 0.25$, the middle of the optimal range, as our mode estimate we find lower bootstrapped standard errors for $h = 0.2$ compared to $h = 0.05$: More observations are included in the mode estimate so that more resampled datasets give the same or similar β estimates. Of course this is an extreme example and for larger datasets the estimated β is not ambiguous for reasonable choices of h, but it nicely highlights the intuition about the trade-off between bias and precision.

To summarize, the algorithm for mode regression is as follows:

1. Choose an initial bandwidth $h > 0$.
2. Choose an initial β, for example from an OLS regression.
3. (a) Calculate residuals $e_i = L_i - x_i'\beta$ for all observations $i = 1 \dots N$.
   (b) Calculate the absolute value of standardized residuals $u_i = |e_i/h|$.

[11] This is close to the optimal bandwidth from kernel estimation as derived below.

(c) Count the number of cases where u_i is larger than 1:

$$Q = \sum_{i=1}^{N} 1\{u_i \geq 1\}.$$

(d) Search for the β that minimizes Q by repeating steps 3(a)-(c).

4. Check how many observations fall outside $2h$ of the mode. If Q/N is more than 40% then increase the bandwidth and go back to step 1. If Q/N is less than 20%, lower the bandwidth and go back to step 1.

5. Calculate standard errors by bootstrap:

   (a) Sample K datasets of the same size as the original data, with replacement.

   (b) for each dataset, $k = 1 \dots K$, compute the β_k vector that minimizes Q, using the same method and bandwidth as in step (3) above.

   (c) Calculate the standard error $\sigma(\widehat{\beta})$ as the standard deviation of β_k across the K sampled datasets.

B Kernel Estimation (NPMLE)

Another method of estimating the mode is to estimate the maximum of the conditional density of L given x. We start with a non-parametric estimate of the joint density of L and x using a *kernel*, $\mathcal{K}_h(u)$:

$$\widehat{f}_h(L, x) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{K}_h(L_i - L)\, \mathcal{K}_h(X_i - x). \tag{28}$$

Two popular choices for the kernel are:

- Epanechnikov: $\mathcal{K}_h(u) = \frac{1}{h} \cdot \frac{3}{4} \left[1 - \left(\frac{u}{h}\right)^2\right] 1\{|\frac{u}{h}| \leq 1\}$.
- Gauss: $\mathcal{K}_h(u) = \frac{1}{h} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(u/h)^2\right)$.

The bandwidth $h > 0$ determines the influence of observations that are further away from x and L. A higher h means more influence of far-away observations (i.e. more smoothing).

Table II

Kernel Mode Estimate with Sample Data, no Covariates

Panel A: Kernel density at each datapoint

L data	0.1	0.2	0.3	0.3	0.3	0.4	0.6	0.7	0.8	0.9
$\mathcal{K}_h(L_i - 0.4)$	0.232	0.963	2.265	2.265	2.265	3.012	0.963	0.232	0.032	0.002

Panel B: Density across leverage ratios

L	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\widehat{f}_h(L,x)$	0.840	1.307	1.480	1.223	0.891	0.816	0.883	0.855	0.648	0.349

The mode estimator is the value of L that maximizes the density:

$$\widehat{m}(x) = \arg\max_L \widehat{f}_h(L,x). \tag{29}$$

Essentially we are doing maximum likelihood but using a non-parametric likelihood, hence the term Non-Parametric Maximum Likelihood Estimation (NPMLE).[12] Romano (1988) derives asymptotic convergence results of this kernel density estimate of the mode.

Coming back to the previous example, we now illustrate how to calculate the kernel density for a guess of the mode of $L = 0.4$ using the Gaussian kernel with a bandwidth $h = 0.1324$. Without a covariate, we can omit the factor $\mathcal{K}_h(X_i - x)$.[13] Table II Panel A gives the kernel density of L, $\mathcal{K}_h(L_i - L)$ for each datapoint. For example, $\mathcal{K}_h(L_2 - 0.4) = \frac{1}{h} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}((0.2 - 0.4)/h)^2\right) = 0.963$. The joint density estimate $\widehat{f}_h(L = 0.4, x) = 1.223$ is simply the average of $\mathcal{K}_h(L_i - L)$ over the ten datapoints.

[12] Note that we want to find the maximum of the *conditional* density $\widehat{f}_h(L|x)$. Maximizing the joint density $\widehat{f}_h(L,x)$ is equivalent since conditioning on x is just scaling the joint density by a constant factor.

[13] With $X = 1$ for all datapoints, $\mathcal{K}_h(0) = 1/(h\sqrt{2\pi})$ for all datapoints. Hence, it is simply a scaling factor and can be omitted for the purpose of finding a mode estimate that maximizes the density. It does matter when we have covariates in X that are not constants, as we illustrate below.

In Table II Panel B we calculate the density estimate for a range of L. The mode estimate is the value of L that maximizes the density. From the table it is clear that the mode estimate is 0.3. As before, using a numerical maximizer is a great alternative to using a grid search as shown here.

With covariates, X, we need to calculate the density at the value of X for which we want to estimate the conditional mode of leverage. Suppose we have one covariate as shown in Table III Panel A. For illustration, let's calculate $\widehat{f}_h(L = 0.4, x = 0.15)$, the density estimate at $L = 0.4$ and $x = 0.15$. Panel A of Table III shows the density $\mathcal{K}_h(X_i - x)$ evaluated at $x = 0.15$. The optimal bandwidth for this X is 0.066 (calculated using the algorithm below). Panel A also shows the product of $\mathcal{K}_h(L_i - L)$ and $\mathcal{K}_h(X_i - x)$ for each datapoint. The joint density estimate $\widehat{f}_h(L = 0.4, x = 0.15) = 5.738$ is simply the average over this product. Repeating this exercise for various values of L, Panel B shows that the joint density is maximized at $L = 0.3$. We conclude that the conditional mode at $x = 0.15$ is $L = 0.3$.

For comparison, Table III also show the calculation of the conditional mode at $x = 0.3$, which turns out to be $L = 0.7$.

The choice of bandwidth is quite important for kernel methods. We use the popular solve-the-equation plug-in method of Sheather-Jones (1991), with a simple adjustment to account for autocorrelation. In theory this bandwidth minimizes the distance between the true and fitted densities. We use a different bandwidth for each variable.[14]

The algorithm for NPMLE estimation of the mode, $\widehat{m}(x)$, for a particular realization of x is:

1. Calculate the optimal bandwidth for leverage, h_L:

    (a) Select a set of observations, $\mathcal{S}$, with observations of each firm that are at least one standard deviation apart (this controls for correlation in residuals between observations that are close, and is only used for calculating

[14] The plug-in method has been shown to have superior convergence rates and to work better in simulations compared to the popular "leave-one-out cross-validation" method (Gasser, Kneip and Köhler (1991)).

Table III

Kernel Mode Estimate with Sample Data, One Covariate X

Panel A: Kernel density at each datapoint for X = 0.15 and X = 0.3

L data	0.1	0.2	0.3	0.3	0.3	0.4	0.6	0.7	0.8
X data	0.05	0.15	0.15	0.15	0.1	0.2	0.35	0.3	0.4
$\mathcal{K}_h(X_i - 0.15)$	1.940	5.991	5.991	5.991	4.519	4.519	0.066	0.474	0.005
$\mathcal{K}_h(L_i - 0.4)\mathcal{K}_h(X_i - 0.15)$	0.449	5.770	13.570	13.570	10.237	13.613	0.064	0.110	0.000
$\mathcal{K}_h(X_i - 0.3)$	0.005	0.474	0.474	0.474	0.066	1.940	4.519	5.991	1.940
$\mathcal{K}_h(L_i - 0.4)\mathcal{K}_h(X_i - 0.3)$	0.001	0.457	1.074	1.074	0.149	5.844	4.353	1.387	0.061

Panel B: Density across leverage ratios for X = 0.15 and X = 0.3

L	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\widehat{f}_h(L, x = 0.15)$	3.635	6.417	7.541	5.738	2.818	0.964	0.317	0.134	0.050	0.012
$\widehat{f}_h(L, x = 0.3)$	0.253	0.576	0.977	1.440	2.195	3.128	3.362	2.490	1.264	0.447

optimal bandwidth).

(b) Calculate the standard deviation, σ, of the selected observations.[15]

(c) Calculate $\Phi_6 = \frac{-15}{16\sqrt{\pi}}\sigma^{-7}$

(d) Calculate $\Phi_8 = \frac{105}{32\sqrt{\pi}}\sigma^{-9}$

(e) Calculate $g_1 = \left[\frac{-6}{\sqrt{2\pi}\Phi_6 S}\right]^{1/7}$, where S is the number of observations in the restricted sample $\mathcal{S}$.

(f) Calculate $g_2 = \left[\frac{30}{\sqrt{2\pi}\Phi_8 S}\right]^{1/9}$

(g) Calculate

$$\Phi_4(g_1) = \frac{1}{S^2}\sum_{i=\mathcal{S}}\sum_{j=\mathcal{S}}\mathcal{K}_{g_1}^{(4)}(L_i - L_j),$$

where $\mathcal{K}_h^{(r)}$ is the r-th derivative of the kernel with bandwidth h.

[15] An alternative to restricting the sample used for calculating the optimal bandwidth is to use the spectral density estimate of the standard deviation at this step to account for correlation between residuals.

(h) Recalculate

$$\Phi_6(g_2) = \frac{1}{S^2} \sum_{i=\mathcal{S}} \sum_{j=\mathcal{S}} \mathcal{K}_{g_2}^{(6)}(L_i - L_j),$$

(i) Find the optimal bandwidth h_L by solving the non-linear equation

$$h_L = \left[\frac{R(\mathcal{K})}{\mu_2(\mathcal{K})^2 \Phi_4[\gamma(h_L)] S} \right]^{1/5},$$

where $\Phi_4[\gamma(h_L)]$ is calculated as in step (g) but using bandwidth $\gamma(h_L)$ instead of g_1:

$$\gamma(h_L) = \left[\frac{-2\mathcal{K}^{(4)}(0)\mu_2(\mathcal{K})\Phi_4(g_1)}{R(\mathcal{K})\Phi_6(g_2)} \right]^{1/7} h_L^{5/7}.$$

The constants $R(\mathcal{K}) = \int \mathcal{K}(x)^2 dx$, $\mu_2(\mathcal{K}) = \int x^2 \mathcal{K}(x) dx$ and $\mathcal{K}^{(4)}(0)$ are determined by the choice of kernel. For example, for the Gaussian kernel $R(\mathcal{K}) = 1/(2\sqrt{\pi})$, $\mu_2(\mathcal{K}) = 1$ and $\mathcal{K}^{(4)}(0) = 3/\sqrt{2\pi}$.

2. Repeat 1(a)-(i) to calculate the optimal bandwidth, h_j, for each covariate x_j, $j = 1 \ldots J$.

3. For each observation $i = 1 \ldots N$, calculate the kernel weight for the covariates:

$$\widehat{f}(X_i) = \prod_{j=1}^{J} \mathcal{K}_{h_j}(X_{ij} - x_j),$$

where X_{ij} is the i-th observation of the j-th covariate and x_j is the j-th component of the vector x for which we want to estimate the mode $\widehat{m}(x)$

4. Calculate the kernel density $f(L, x)$ for a grid of $L = 0 \ldots 1$:

$$\widehat{f}(L, x) = \sum_{i=1}^{N} \left[\mathcal{K}_{h_L}(L_i - L) \cdot \widehat{f}(X_i) \right],$$

where L_i is the leverage of the i-th observation.

5. Find the L in the grid that maximizes $\widehat{f}(L, x)$. This is the point estimate of the mode, $\widehat{m}(x)$.

6. Calculate standard errors by bootstrap:

(a) Sample K datasets of the same size as the original data, with replacement.

(b) for each dataset, $k = 1 \dots K$, repeat steps 3 through 5 to compute the mode estimate $\widehat{m}_k(x)$.

(c) Calculate the standard error $\sigma(\widehat{m}(x))$ as the standard deviation of $\widehat{m}_k(x)$ across the K sampled datasets.

To calculate the another point estimate and standard errors at a different x, we can skip the bandwidth estimation step and start at 3(a).

Finding the optimal bandwidth is a bit involved. We show its implementation for our ten datapoint sample of the leverage variable, using all ten datapoints, for illustration. First, the standard deviation of L is 0.272. This gives the values $\Phi_6 = -4849.1$, $\Phi_8 = 23004$, $g_1 = 0.243$ and $g_2 = 0.259$. For the next step we need the fourth and sixth derivative of the Gaussian kernel:

$$\mathcal{K}_h^{(4)}(u) = \frac{1}{h^5\sqrt{2\pi}} \exp\left(-\frac{1}{2}(u/h)^2\right) \cdot \left[3 - 6(u/h)^2 + (u/h)^4\right] \tag{30}$$

$$\mathcal{K}_h^{(6)}(u) = \frac{1}{h^7\sqrt{2\pi}} \exp\left(-\frac{1}{2}(u/h)^2\right) \cdot \left[-15 + 45(u/h)^2 - 15(u/h)^4 + (u/h)^6\right] \tag{31}$$

We calculate $\Phi_4(g_1)$ by summing $\mathcal{K}_{g_1}^{(4)}(L_i - L_j)$ for all combinations of i and j and find $\Phi_4(g_1) = 187.226$. Similarly we calculate $\Phi_6(g_2) = -9952.7$, using the sixth derivative of $\mathcal{K}$ and the bandwidth g_2. Now we iterate on $\gamma(h_L)$ and h_L in step 1(i) until we find the bandwidth h_L that solves both equations. We start with a bandwidth $h_L = g_1 = 0.243$. This yields a value of

$$\gamma(h_L) = \left[\frac{-2 \cdot (3/\sqrt{2\pi}) \cdot 187.226}{1/(2\sqrt{\pi}) \cdot -9952.7}\right]^{1/7} 0.243^{5/7} = 0.280. \tag{32}$$

We recalculate $\Phi_4(\gamma(h_L)) = 93.630$ and plug in to find

$$h_L = \left[\frac{1/(2\sqrt{\pi})}{93.630}\right]^{1/5} = 0.198. \tag{33}$$

Iterating on the latter two equations converges to an optimal bandwidth of $h_L = 0.1324$.

C Comparison of the Methods and Simulation

Mode regression is closely related to kernel estimation. Essentially, we can think of mode regression as kernel estimation with the kernel $1\left\{\left|\frac{L_i - m(x)}{h}\right| \leq 1\right\}$ for the dependent, and a flat kernel for the conditioning variables. One important difference is that the mode regression assumes that the mode is a globally linear function of x, whereas kernel estimation presumes it is locally constant. Although this makes the mode regression estimates easier to compare to OLS estimates, the linear form is restrictive. Moreover, the asymptotic and optimal bandwidth properties of kernel estimators are better developed.

To get a sense of the empirical performance of the mode regression in comparison to a mean estimator such as OLS, we run a set of simulations on our simple dynamic model of capital structure presented above. Specifically, for each firm we draw a random coupon from a uniform distribution between 0.12 and 0.9, and a random refinancing boundary, δ_R, from a uniform distribution between 1.1 and 2.5. The other parameters are as before. Each firm now has a different target leverage and refinancing boundary. Holding the parameters constant over time for each firm, we simulate 10,000 firms for 10 years.

If we could observe the true L^* in the final year and regress the cross-section of target leverage ratios on the exogenous coupon and δ_R, then we would find the results in the left-most column of the table below. However, in reality we do not observe L^* but instead we have to work with L. Using OLS regression on the final cross-section of leverage, the middle column of the table reveals that the coefficients on both coupon and δ_R are severely biased downwards, and the intercept is biased upwards. In contrast, the mode regression estimates are very close to their true values.

We conclude that the conditional mode estimates do a good job of identifying the coefficients in cross-sectional capital structure regressions where moment estimators fail.

Table IV

	True	OLS	Mode regression
Intercept	0.051	0.428	0.054
coupon	0.564	0.379	0.569
δ_R	0.013	-0.047	0.009

III Data

We collect panel data on leverage and company characteristics from Compustat between 1950 and 2005. Leverage (L) is measured as the book value of debt divided by the book value of debt plus the market value of equity. Many firms persistently have zero leverage. To avoid the issue of bimodality of the leverage distribution, we drop all firm-years with leverage smaller than 0.05 (see Strebulaev and Yang (2008)). Profitability ($PROF$) is defined as EBITDA over sales, truncated at ±100%. Depreciation relative to book assets ($DEPR$) is truncated between 0 and 1. We define tangible assets (PPE) as property, plant & equipment divided by book assets. Firm size ($SIZE$) is the natural logarithm of book assets, and M/B is the equity market-to-book ratio, truncated between 0 and 10. The final sample contains 137,623 firm-years.

Table **??** reports summary statistics. Mean (median) leverage is 0.371 (0.323). This is somewhat high relative to the extant literature because we eliminated the extremely low-leverage firms. In contrast, the mode of leverage is 0.250.

IV Cross-sectional Results

TO BE COMPLETED

V Conclusion

Traditional moment-based estimators are flawed measures of target leverage in dynamic models of capital structure with fixed adjustment costs. We establish that the conditional mode is a theoretically motivated and robust measure of target capital structure, and we introduce empirical methods to estimate the conditional mode of leverage.

We present two main empirical results: i) optimal leverage is lower than traditionally thought, and; ii) asset tangibility is a relatively more important determinant of capital structure than market-to-book ratio, compared to results in the literature. The latter finding suggests that collateralization is deserving of more attention, whereas theories of debt overhang may be less important in explaining capital structures.

Our methodology needs to be subjected to more robustness tests. Most importantly, the robustness of our results to incomplete refinancing when firms opportunistically rebalance (e.g. a maturing debt issue) needs to be explored. In addition, the impact of systematic shocks to firms' leverage ratios should be investigated. Finally, relaxing the stationarity assumption.

On the methodological side, an interesting avenue for future research is to filter the estimates of the conditional mode. This would allow the target leverage of the firm to vary over time. Filtering estimates of the conditional mode are used in meteorology and imaging applications (known as the "variational approach").

References

Byoun, Soku, 2008, How and when do firms adjust their capital structures towards targets?, *Journal of Finance* 63, 3069-3096.

Chang, Xin and Sudipto Dasgupta, 2009, Target behavior and financing, *Journal of Finance* 64, 1767-1796.

Dangl and Zechner, 2003, Credit risk and dynamic capital structure choice, *Journal of Financial Intermediation* x, x-x.

Deschênes, Olivier, and Michael Greenstone, 2007, The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather, *American Economic Review* 97, 354-385.

Fischer, Edwin O., Robert Heinkel, and Josef Zechner, 1989, Dynamic capital structure choice: Theory and tests, *Journal of Finance* 44, 19-40.

Flannery, Mark J., and Kasturi P. Rangan, 2006, Partial adjustment toward target capital structures, *Journal of Financial Economics* 79, 469-506.

Gasser, Theo, Alois Kneip and Walter Köhler, 1991, A Flexible and Fast Method for Automatic Smoothing , *Journal of the American Statistical Association* 86, 643-652.

Haerdle, Wolfgang, 1995, Applied nonparametric regression, Cambridge University Press.

Harris, Milton, and Arthur Raviv, 1991, The theory of capital structure, *Journal of Finance* 46, 297-355.

Herrman, Gasser and Kneip, 1992, Choice of bandwidth for kernel regression when residuals are correlated, *Biometrika* 79, 783-795.

Huang, Rongbing and Jay Ritter, 2009, Testing theories of optimal capital structure and estimating the speed of adjustment, *Working Paper*, University of Florida.

Iliev, Peter, and Ivo Welch, 2009, How quickly do firms readjust capital structure, *Working Paper*, Penn State and Brown University.

Karlin, Samuel, and Howard M. Taylor, 1981, A second course in stochastic processes, Academic Press: New York.

Kayhan, Ayla, and Sheridan Titman, 2007, Firms' histories and their capital structures, *Journal of Financial Economics* 83, 1-32.

Korajczyk, Robert A., and Amnon Levy, 2003, Capital structure choice: Macroeconomic conditions and financial constraints, *Journal of Financial Economics* 68, 75-109.

Korteweg, Arthur G., 2010, The net benefits to leverage, *Working Paper*, Journal of Finance, forthcoming.

Lee, Myoung-Jae, 1989, Mode regression, *Journal of Econometrics* 42, 337-349.

Lee, Myoung-Jae, 1993, Quadratic mode regression, *Journal of Econometrics* 57, 1-19.

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

Romano, Joseph P., 1988, On Weak Convergence and Optimality of Kernel Density Estimates of the Mode, *The Annals of Statistics* 16, 629-647.

Strebulaev, Ilya A., 2007, Do tests of capital structure theory mean what they say?, *Journal of Finance* 62, 1747–1787.

Strömberg, David, 2004, Radio's Impact on Public Spending, *Quarterly Journal of Economics* 119, 189-221.

Titman, Sheridan and Roberto Wessels, 1988, The determinants of capital structure choice, *Journal of Finance* 43, 1-19.

Appendix: Proofs

In this appendix we derive the stationary distribution of leverage, conditions under which the mode of this distribution is the optimal leverage ratio upon refinancing. We close with some examples, in particular the geometric Brownian Motion used in the main text.

The stationary distribution of leverage.

Assume that between refinancing points, leverage (L_t) follows a stochastic process:

$$dL_t = \mu(L_t)dt + \sigma(L_t)dB_t \tag{34}$$

with a lower and upper resetting barrier at L^d and L^u, respectively. When the firm hits either a lower resetting barrier, L^d, or an upper barrier L^u then L_t is reset to the optimal ratio L^*.

We derive the form of the stationary distribution and the necessary and sufficient condition for its existence.

If it exists, the stationary distribution of leverage, $f(L)$, satisfies the well-known forward equation (e.g. Karlin and Taylor, 1981, p.220-221):

$$\frac{1}{2}\frac{\partial^2}{\partial L^2}\left(\sigma^2(L)f(L)\right) - \frac{\partial}{\partial L}\left(\mu(L)f(L)\right) = 0. \tag{35}$$

To find $f(L)$, first integrate the forward equation with respect to L:

$$\frac{\partial}{\partial L}\left[\sigma^2(L)f(L)\right] - 2\mu(L)f(L) = A_1, \tag{36}$$

with A_1 the constant of integration.

Define

$$\gamma(L) \equiv \exp\left[-\int^L \frac{2\mu(y)}{\sigma^2(y)}dy\right], \tag{37}$$

and multiply (36) by (37):

$$\frac{\partial}{\partial L}\left(\gamma(L)\sigma^2(L)f(L)\right) = A_1\gamma(L). \tag{38}$$

Integrate with respect to L to recover the stationary distribution:

$$f(L) = m(L)\left[A_1\int^L \gamma(y)dy + B_1\right], \tag{39}$$

where B_1 is another constant of integration, and

$$m(L) \equiv \frac{1}{\gamma(L)\sigma^2(L)}. \tag{40}$$

We postulate the solution:

$$f(L) = m(L)\left[A_1 \int^L \gamma(y)dy + B_1\right] \quad \text{for} \quad L^d \leq L \leq L^*, \tag{41}$$

$$f(L) = m(L)\left[A_2 \int^L \gamma(y)dy + B_2\right] \quad \text{for} \quad L^* \leq L \leq L^u, \tag{42}$$

and impose the following four boundary conditions to pin down the four constants A_1, A_2, B_1 and B_2:

1. $f(L^d) = 0$,

2. $f(L^u) = 0$,

3. $f(L^*-) = f(L^*+)$,

4. $\int_{L^d}^{L^u} f(L)dL = 1$.

The first two conditions imply

$$f(L) = A_1 m(L)\Gamma(L^d, L) \quad \text{for} \quad L^d \leq L \leq L^*, \tag{43}$$

and

$$f(L) = -A_2 m(L)\Gamma(L, L^u) \quad \text{for} \quad L^* \leq L \leq L^u, \tag{44}$$

with the standard scale measure defined as

$$\Gamma(a, b) \equiv \int_a^b \gamma(y)dy. \tag{45}$$

To ensure non-negativity of the pdf, $A_1 > 0$ and $A_2 < 0$.

The third condition enforces a continuous distribution function and results in

$$A_1\Gamma(L^d, L^*) = -A_2\Gamma(L^*, L^u). \tag{46}$$

The fourth condition ensures that we have a proper probability distribution function and requires

$$A_1 \int_{L^d}^{L^*} m(L)\Gamma(L^d, L)dL - A_2 \int_{L^*}^{L^u} m(L)\Gamma(L, L^u)dL = 1 \tag{47}$$

We group the latter two conditions into a linear system:

$$X \cdot \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{48}$$

where

$$X = \begin{bmatrix} \Gamma(L^d, L^*) & \Gamma(L^*, L^u) \\ \int_{L^d}^{L^*} m(L)\Gamma(L^d, L)dL & -\int_{L^*}^{L^u} m(L)\Gamma(L, L^u)dL \end{bmatrix} \tag{49}$$

A stationary distribution exists *if and only if* the determinant of the matrix X is non-zero.

The mode of $f(L)$

Assuming that a stationary leverage distribution exists, we show that L^* is the mode of the stationary distribution under the following sufficient conditions on $\mu(L)$, $\sigma(L)$, L^d, L^* and L^u:

$$\mu(L) - \sigma(L) \cdot \sigma'(L) > -\frac{1}{2}\frac{1}{\Gamma(L^d, L)m(L)} \quad \text{for} \quad L^d < L < L^* \tag{50}$$

$$\mu(L) - \sigma(L) \cdot \sigma'(L) < \frac{1}{2}\frac{1}{\Gamma(L, L^u)m(L)} \quad \text{for} \quad L^* < L < L^u \tag{51}$$

PROOF: Sufficient conditions under which L^* is the mode are (note that $f(L)$ is not necessarily differentiable at L^*):

1. $f'(L) > 0$ for $L^d < L < L^*$,

2. $f'(L) < 0$ for $L^* < L < L^u$.

For $L^d < L < L^*$, take the derivative of (43) with respect to L,

$$f'(L) = \frac{m'(L)}{m(L)} f(L) + \frac{A_1}{\sigma^2(L)}. \tag{52}$$

Plugging in

$$\frac{m'(L)}{m(L)} = \frac{2\mu(L)}{\sigma^2(L)} - \frac{2\sigma'(L)}{\sigma(L)}, \tag{53}$$

and equation (43) yields

$$f'(L) = A_1 \left(\frac{2\mu(L)}{\sigma^2(L)} - \frac{2\sigma'(L)}{\sigma(L)} \right) m(L)\Gamma(L^d, L) + \frac{A_1}{\sigma^2(L)}. \tag{54}$$

We know that $A_1 > 0$, so $f'(L) > 0$ iff

$$\frac{\mu(L) - \sigma(L)\sigma'(L)}{\sigma^2(L)} > -\frac{1}{2}\frac{\gamma(L)}{\Gamma(L^d, L)}. \tag{55}$$

Similarly, condition (51) follows from the derivative of equation (44) and $A_2 < 0$. The conditions are not dependent on the values of A_1 and A_2, but depend only on the specification of $\mu(L)$ and $\sigma(L)$.

Examples

1. Brownian Motion

Consider the standard Brownian Motion with constant drift and variance i.e. $\mu(L_t) = \mu$ and $\sigma(L_t) = \sigma$. Therefore

$$\gamma(L) = \exp\left[-\int^L \frac{2\mu(y)}{\sigma^2(y)}dy\right] = \exp\left(-kL\right), \tag{56}$$

$$\Gamma(a,b) = \int_a^b \gamma(y)dy = \frac{1}{k}\left[\exp(-ka) - \exp(-kb)\right], \tag{57}$$

where $k \equiv 2\mu/\sigma^2$.

Next we solve (48) for A_1 and A_2:

$$A_1 = -k\sigma^2 / \left[-(L^d - L^*) + (L^u - L^*)\tfrac{1-\gamma(L^d-L^*)}{1-\gamma(L^u-L^*)}\right] \tag{58}$$

$$A_2 = -k\sigma^2 / \left[-(L^d - L^*)\tfrac{1-\gamma(L^u-L^*)}{1-\gamma(L^d-L^*)+(L^u-L^*)}\right]. \tag{59}$$

Plugging into (43) and (44) yields the stationary distribution of L:

$$f(L) = \frac{1-\gamma(L^d-L)}{-(L^d-L^*)+(L^u-L^*)\frac{1-\gamma(L^d-L^*)}{1-\gamma(L^u-L^*)}} \quad \text{for} \quad L^d < L \leq L^* \tag{60}$$

$$f(L) = \frac{1-\gamma(L^u-L)}{-(L^d-L^*)\frac{1-\gamma(L^u-L^*)}{1-\gamma(L^d-L^*)}+(L^u-L^*)} \quad \text{for} \quad L^* < L < L^u. \tag{61}$$

For $k \neq 0$, this stationary distribution exists only if $L^d < L^* < L^u$.

Typically, $k < 0$ since the expected return to equity is higher than debt. Conditions (50) and (51) under which the mode of the distribution is L^*, are then

$$\frac{1}{1-\gamma(L^d - L)} > 1 \quad \text{for} \quad L^d < L < L^* \tag{62}$$

$$\frac{1}{1-\gamma(L^u - L)} < 1 \quad \text{for} \quad L^* < L < L^u, \tag{63}$$

Both conditions clearly hold for $k < 0$.

2. Geometric Brownian Motion

The drift and volatility of the geometric Brownian motion are $\mu(L_t) = \mu L_t$ and $\sigma(L_t) = \sigma L_t$. By Itô's lemma this is equivalent to the natural logarithm of L_t following a Brownian motion with drift $\mu - \frac{1}{2}\sigma^2$:

$$d\log(L_t) = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dB_t \tag{64}$$

Since the geometric Brownian motion is a strictly increasing transformation of a standard Brownian motion, the results for existence of a stationary distribution and the mode of the distribution that we derived for the standard Brownian motion apply to the geometric Brownian motion as well.

By changing variables from $\log(L)$ to L we find the expression for the stationary distribution of L:

$$f(L) = \left[\frac{L^d}{L} - \left(\frac{L^d}{L}\right)^{-k+2}\right]/C_1 \quad \text{for} \quad L^d < L \leq L^* \tag{65}$$

$$f(L) = \left[\frac{L^u}{L} - \left(\frac{L^u}{L}\right)^{-k+2}\right]/C_2 \quad \text{for} \quad L^* < L < L^u, \tag{66}$$

for $k \neq 0$, where $k \equiv 2\mu/\sigma^2$ and

$$C_1 \equiv L^d\left[-\log\left(\frac{L^d}{L^*}\right) + \log\left(\frac{L^u}{L^*}\right)\cdot\frac{1-(\frac{L^d}{L^*})^{-k+1}}{1-(\frac{L^u}{L^*})^{-k+1}}\right] \tag{67}$$

$$C_2 \equiv L^u\left[-\log\left(\frac{L^d}{L^*}\right)\cdot\frac{1-(\frac{L^u}{L^*})^{-k+1}}{1-(\frac{L^d}{L^*})^{-k+1}} + \log\left(\frac{L^u}{L^*}\right)\right]. \tag{68}$$

The mode of this distribution is L^* iff $L^* < \left(\frac{1}{2-k}\right)^{\frac{1}{1-k}} L^d$ and $k < 2$. The mean is

$$\begin{aligned} E(L) = {} & \frac{L^d}{C_1}\left(L^* - L^d\right) + \frac{L^u}{C_2}\left(L^u - L^*\right) + \\ & \frac{(L^d)^2}{kC_1}\left[1 - \left(\frac{L^d}{L^*}\right)^{-k}\right] - \frac{(L^u)^2}{kC_2}\left[1 - \left(\frac{L^u}{L^*}\right)^{-k}\right]. \end{aligned} \tag{69}$$