Dynamic Models and Structural Estimation in Corporate Finance

Ilya A. Strebulaev and Toni M. Whited*

first version, June 2012
revised, December 2012

ABSTRACT

We review the last two decades of research in dynamic corporate finance, focusing on capital structure and the financing of investment. We first cover continuous time contingent claims models, starting with real options models, and working through static and dynamic capital structure models. We then move on to corporate financing models based on discrete-time dynamic investment problems. We cover the basic model with no financing, as well as more elaborate models that include features such as costly external finance, cash holding, and both safe and risky debt. For all the models, we offer a minimalist, simplified presentation with a great deal of intuition. Throughout, we show how these models can answer questions concerning the effects of financial constraints on investment, the level of corporate leverage, the speed of adjustment of leverage to its target, and market timing, among others. Finally, we review and explain structural estimation of corporate finance models.

*Strebulaev is from Stanford Graduate School of Business and the National Bureau of Economic Research; Whited is from the University of Rochester. Yunjeen Kim provided excellent research assistance. We are grateful for helpful discussions with Michael Brennan, Harry DeAngelo, Will Gornall, Dirk Hackbarth, Akitada Kasahara, Arthur Korteweg, Boris Nikolov, Yuri Salitsky, Amit Seru, and Luke Taylor. We also would like to thank the anonymous reviewer and Franklin Allen (the Editor) for helpful suggestions.

Contents

1 Introduction

Over the last 20 years, research on dynamic corporate finance has witnessed dramatic growth in both theoretical and empirical directions. Several interrelated factors have led to this advancement. First, although it has long been recognized that most financial decisions involve inherently dynamic interactions (e.g. Lintner 1956), the development of methods necessary to tackle a number of important dynamic issues has lagged. However, advances in stochastic dynamic optimization techniques, contingent claims asset pricing analysis, game theory, and dynamic investment modeling have opened avenues for the fruitful study of dynamic problems in corporate finance. Second, there has been contemporaneous development in structural empirical methods. Third, researchers have gained access to more and significantly higher-quality data, as well as to dramatically better computing power, which makes analyzing such data possible in a reasonable amount of time. These resources have allowed researchers to take dynamic models and empirical methods seriously and, more importantly, to ask more demanding questions from the data and methods. Fourth, it is gradually becoming clearer that static models and the intuition they imply often fail to explain even simple, first-order stylized facts. In contrast, a dynamic paradigm allows the formalization and exploration of new questions that are either irrelevant or impossible to address in a traditional set-up.

Given this background, the goal of this survey is threefold. First, we wish to explain the models and techniques used in this literature as simply as possible, with the goal of making the literature more accessible. Many of the published papers in dynamic corporate finance contain models with many details. Although detail adds to the realism and rigor of the research, the unintended consequence is that the models appear to be black boxes with many indistinguishable moving internal parts. In reality, the intuition behind the models used in this literature is simple, and our goal is to reveal this simplicity. As such, we do not claim to offer a full-blown and complete overview of dynamic theoretical and empirical models in dynamic corporate finance. We also do not claim to be rigorous, precise, or generic. Rather, we offer an intuitive presentation with much less mathematical formalism than has been used in the papers we survey. For technical details and formal proofs of many results, we refer readers to the appropriate studies.

Our second goal is to introduce the reader to the main strands of this literature. Thus, our paper can be viewed in part as a literature review and in part as a tutorial. We spend a great deal of time on dynamic capital structure and on corporate investment, both of which have been the home of much, though certainly not all, of the work in dynamic corporate finance. Looking at these areas via the lens of a dynamic model helps us understand questions that would be impossible to tackle in a static framework. For example, the optimal *timing* of investment projects, equity issuance, and debt refinancing is inherently dynamic, as is the *speed* of adjustment of a leverage ratio to its target. Taking dynamics seriously also helps shed new light on questions that can be examined in static frameworks. Important examples include the detection of financial constraints, the corporate diversification discount, the low-leverage puzzle, the effects of agency on leverage and investment, and the anomalous negative sensitivity of leverage to income, which stands in the face of static capital structure theories that predict higher leverage for high-income firms.

One particular advantage of using dynamic models is that they can often provide *quantitative*, rather than simply qualitative implications. This distinction is especially important in areas such as capital structure, in which the relative magnitudes of the costs and benefits of leverage have been the center of much of the research agenda. Throughout we delineate the similarities and differences between the different classes of dynamic models that we review.

Our third goal is to explain how to estimate dynamic models. Once again, our intent is not to provide a rigorous econometric treatment but a practical, "hands-on" guide to three specific methodologies that have been used in the literature: generalized method of moments, simulated method of moments, and maximum simulated likelihood. We also provide a concise guide to the extant structural estimation literature in corporate finance.

Space constraints necessitate several compromises. We do not cover recent developments in dynamic principal agent models. We also omit discussion of the relatively widely used class of dynamic models based on two to three time periods. We also leave out the large, closely related literature on dynamic models of credit spreads.

The rest of the paper proceeds as follows. Section 2 provides an overview of dynamic corporate finance models based on techniques developed in the continuous time contingent claims literature. We start with real options and then move on to static and dynamic capital structure models. We conclude this section by surveying the rest of the literature based on this class of models. Section 3 covers a separate strand of the literature that stems from a completely different formal base—discrete time investment models. Here, we cover the basic model without financing, and then move through the literature by expanding this basic model one feature at a time. Again, we conclude this section with a broad overview of the literature. Section 4 then reviews the relatively small number of different econometric techniques that have been used to estimate these models, as well as the studies that have used them. We close with a brief overview of directions for future research. Because we are reviewing highly diverse sets of models, we need to use a great deal of notation. We therefore define all of the symbols that we use below in Table 1.

2 Dynamic Contingent Claims Models

We start by providing an informal, intuitive discussion that illuminates the nature of dynamic contingent claims models and their contribution to our understanding of corporate finance. We then present the basic structure of several concrete models that serve as fundamental building blocks and discuss useful implications and applications.

2.1 Introduction

Research that explores dynamic contingent claims models[1] sits at the intersection of asset pricing and corporate finance. It has been subject to particularly rapid development in recent years. These models borrow from extensive theoretical developments in related fields, such as asset pricing, macroeconomics, and stochastic processes.

To understand how these models work generically and how they differ from other theoretical frameworks in corporate finance, note that they start with the acknowledgment that any claims

[1]These models are often confusingly called structural dynamic models, mostly for historical reasons rather than for any connection to other uses of "structural" in economics.

on corporate cash flow streams are derivatives on underlying firm value or firm cash flows. This means that we can apply option pricing methods to value these claims, assuming initially that all financial and investment decisions by economic agents are fixed. Indeed, the efforts by Black and Scholes (1973) were originally aimed at describing the pricing of corporate securities such as warrants rather than exchange-traded options. In his famous quote, Merton (1974) asserts that "while options are highly specialized and relatively unimportant financial instruments ... the same basic approach could be applied in developing a pricing theory for corporate liabilities in general" (p. 449).

The exact nature of the underlying variables depends on the specific application (e.g., stock price in the Black-Scholes model; firm value, firm cash flows, or prices of firm output and input in other corporate finance settings). These underlying variables are usually the primitives of the model in the sense that their evolution is (at least partially) exogenous to decision making by economic agents, and they constitute so-called state variables.[2] To value corporate securities, the model typically requires the same inputs as a standard option pricing model, including a law for the time variation of state variables that accounts for the distribution of future cash flows. (In the Black-Scholes model, it is a geometric Brownian motion process for the stock price.)

It is useful to relate this set-up to its counterpart in a typical static model. In the latter, the distribution of a benchmark one-period return on an investment project is typically given exogenously; as a natural "state-of-the-world" variable, it thus serves a purpose similar to a law for the time variation of an exogenous state variable in a dynamic framework. Most of the inputs are also not unique to a dynamic problem; for example, there is a well-known relation between an essentially one-period binomial option pricing model and the continuous-time Black-Scholes model, with a tight link between the parameters that appear in these setups.

The inherently dynamic nature of the state variable process in the Black-Scholes-Merton framework (in the original setup, stock price) is an obvious reason for including "dynamic" in the name of these models. More substantial grounds for such a name include the flexibility of this framework,

[2]Note that state variables do not need to be firm-specific. For example, stochastic interest rates are not.

which enables us to study the *interrelation of agents' decisions at different points in time*, as will become clearer in our subsequent discussion.

A second critical element of these models is the objective function of a single decision maker (e.g., shareholders, managers, central planners) or many decision makers (e.g., principals and agents, oligopolists, price-takers in a competitive industry). This feature represents a crucial departure from derivatives models, though not necessarily from asset pricing models at large. An objective function summarizes the desires of the economic agent(s) in question and can be expressed succinctly by combining certain contingent claims. For example, shareholders want to maximize the value of their equity, whereas a benevolent central planner has the value-weighted utility of all the agents in mind. Because the model is dynamic, the objective function can change subtly over time to reflect the changing economic environment (a feature commonly omitted in one-period corporate finance models but discussed at length in literature on contracting). For example, these models can exploit differences between ex ante and ex post objectives of decision makers and naturally give rise to realistic conflicts of interest. They also allow the decision maker's identity to change over time, as a function of the state variables' evolution. For example, equityholders decide on the original firm's financial structure, but debtholders can decide on the course of actions in the event of a default.

A third component is the specification of the set of instruments, known as *controls*, that are available for the decision maker to maximize the objective function, as well as the set of constraints imposed on these controls. For example, equityholders maximize equity value by choosing the firm's current and future investment and financial policies, subject to a limited liability condition. The precise set of control(s) depends on the nature of a particular application. In one of the models we consider in detail later, equityholders choose the amount of debt first, and the timing of default later.

Finally, the model is solved by finding the optimal set of (generally, time-varying) controls at the decision maker's disposal to maximize her objective function. Most technical complications arise at this stage, because, in most realistic specifications, it is notoriously difficult to find a closed-form (and sometimes even approximate) solution. Furthermore, the trade-off between complexity and realism in the model and the feasibility of technical methods grows increasingly challenging. Fortunately though, a dramatic increase in our abilities to solve such models has emerged from the development of (now well-understood) techniques, which can deal with various technical issues, in fields such as stochastic dynamic programming. For example, the smooth-pasting and super-contact conditions (see, among others, Dumas 1991; Dixit and Pindyck 1994; Stokey 2009) have dramatically facilitated the applicability of such models. In particular, these techniques allow researchers to characterize the properties of an optimal solution, even though an explicit solution may not be available.

As should be clear by now, this approach to modeling is very similar in its fundamental qualities to traditional corporate finance models that specify objective functions, choice variables, and actions of economic agents in a similar fashion. They also often ask a very similar set of questions. In fact, any traditional corporate finance model can be reformulated using the language and apparatus of a contingent claims model. However, whether a resulting model is easy to solve is another matter.

A key difference from the paradigm of traditional corporate finance models, whether they are static or dynamic in nature, should also be clear at this point. In traditional models, the values underlying an objective function are separated from any asset pricing implications so that they are alienated from the prices established in markets. This separation has significant consequences, because even if a traditional model can produce a directional prediction of the agent's response to exogenous changes in primitive parameters, not many people would seriously attempt to claim that a traditional model can make effective quantitative predictions. By virtue of their tight link to generic option pricing models, whose primary goal is to price securities, dynamic contingent claims models can generate internally consistent testable quantitative predictions about the behavior of the model's output variables.

Whether quantitative predictions are of interest depends, of course, on the research objective. However, the dynamic approach can also deliver qualitative results and predictions that are economically quite different and sometimes completely absent from the traditional paradigm. In effect, the dynamic paradigm broadens our toolkit and enables us to ask new questions, as well as identify and model new economic forces. In what follows, we illustrate both the quantitative and qualitative sides of dynamic contingent claims models using concrete applications.

2.2 Real Options and Dynamic Investment

One of the first applications of option pricing models to economic decisions gave rise to a new research area known as real options. An important component of the dynamic contingent claims paradigm, the real options framework originally referred to activities in the "real" brick-and-mortar or non-financial world, such as corporate investment policies (hence the name "real"). Research by McDonald and Siegel (1985, 1986) and Brennan and Schwartz (1985) pioneered the field. As an example of a real option model, as well as our first contingent claims setup, we consider a modified version of the investment problem studied by McDonald and Siegel (1985, 1986). As we proceed, we also informally introduce necessary technical machinery.

2.2.1 Real Option Model

The underlying state variable is the value of an investment project that could be undertaken by a firm at a certain fixed cost. If undertaken, this investment decision is irreversible (i.e., the fixed cost is non-recoverable). McDonald and Siegel assume that the state variable follows the same dynamic process as the underlying state variable in the Black-Scholes model (i.e., the stock price on which the option is written).

In a traditional corporate finance paradigm (which remains central in any MBA-level finance textbook, e.g. see Berk and DeMarzo (2011)), an investment decision depends on the net present value (NPV) of the project. McDonald and Siegel introduce a dynamic element in this decision-making process. Instead of a take-it-or-leave-it investment decision that characterizes the NPV decision process, the firm has the flexibility to wait before investing. The (real) option of waiting to invest is valuable, because the future will bring partial resolution of uncertainty about the value of the project. Crucially, the firm has control over the decision of *when* to invest; in other words, the firm decides the optimal time of investment. The value of the option stems from the firm's ability to mitigate the consequences of unfortunate scenarios in which bad news occurs shortly after the firm undertakes what previously appeared to be a positive NPV project. The firm faces a trade-off between delaying investment to minimize the impact of bad news and missing out on the present value of project's cash flows in the meantime. Intuitively, as the project's cash flow level grows, the option to wait declines in value, and it becomes more attractive to invest today. The firm invests when the cash flow level reaches a certain upper threshold.

To develop this model formally, we assume that a decision maker (e.g., firm's manager) continuously observes a potential project's cash flow X that follows, as in Black and Scholes (1973), a geometric Brownian motion process:

$$dX_t = \mu X_t dt + \sigma X_t dW_t^{\mathbf{Q}}, \tag{2.1}$$

where μ and σ are the constant instantaneous growth rate and volatility of cash flow, respectively, and $dW_t^{\mathbf{Q}}$ is a standard Brownian motion process under the risk-neutral measure.[3] The cost of investment is fixed at a constant I. After the investment is made, the firm perpetually generates cash flow X, whereas prior to the investment no cash flow is generated. The manager's control variable is the decision about when to invest in the project. Thus framed, the problem refers to the choice of the time of investment. However, X is the only state variable, so the manager effectively chooses the cash flow level at which it is optimal to invest. We denote this level X_I.

Then the value of the firm can be written as:

$$V(X_0) = A(X_0, X_I)\left(\frac{X_I}{r-\mu} - I\right), \tag{2.2}$$

where $A(X_0, X_I)$ is the date-0 value of a contingent claims security that pays \$1 when the cash flow level X reaches the level X_I for the first time from below, and 0 otherwise. Security A is an example of an Arrow-Debreu-like claim, which is fundamental in any contingent claims model. If and when the threshold level of cash flow is reached for the first time, the investment takes place,

[3] Mapping between physical and risk-neutral measures is economically non-trivial in the case of real assets. In general, decision makers demand a risk premium to the extent that the project's risk is systematic, which in turn exerts an impact on the level of the growth rate μ. All the models we develop, though, can be extended to hold for a generic risk premium. Alternatively, we might assume that the decision maker and the market are risk-neutral or that a project's risk is uncorrelated with systematic risk.

and the firm incurs a cost of I. In return, the firm perpetually receives the project's cash flow, which equals $\frac{X_I}{r-\mu}$ at that date, where r is the constant risk-free rate.

To relate the decision-making process of the manager to a traditional setup, we denote the NPV of the project as $V^{NPV}(X_I) = \left(\frac{X_I}{r-\mu} - I\right)$. We can thus rewrite the value in Equation (2.2) as:

$$V(X_0) = A(X_0, X_I)V^{NPV}(X_I). \tag{2.3}$$

This formulation reveals that the difference created by flexibility is captured entirely by the contingent claim A, so its properties are of great interest to us. Because X follows a geometric Brownian motion, there is a simple closed-form expression for $A(X_0, X_I)$:

$$A(X_0, X_I) = \begin{cases} \left(\frac{X_0}{X_I}\right)^{\xi_2}, & X_0 < X_I; \\ 1, & X_0 \geq X_I, \end{cases} \tag{2.4}$$

where ξ_2 is a function of the constant primitive parameters μ, σ, and r. It is the positive root of the fundamental quadratic equation and is given by:

$$\xi_2 \quad = \quad -\frac{1}{\sigma^2}\left(\mu - \frac{\sigma^2}{2} - \sqrt{\left(\mu - \frac{\sigma^2}{2}\right)^2 + 2\sigma^2 r}\right). \tag{2.5}$$

2.2.2 Optimal Investment Policy

If $X_0 \geq X_I$, the investment occurs instantaneously. In this case, we revert to a traditional NPV analysis. If $X_0 < X_I$, the first-order condition leads to[4]

$$X_I = \frac{\xi_2}{\xi_2 - 1} I(r - \mu). \tag{2.6}$$

Consider the more interesting case of $X_0 < X_I$. The traditional NPV rule prescribes an investment if $\frac{X_I}{r-\mu} > I$, which implies that the investment threshold is $X_I^{NPV} = I(r - \mu)$. It is easy to show that $\xi_2 > 1$, and thus, $X_I > X_I^{NPV}$. This fundamental result underscores the dramatic

[4] It is straightforward to check that the second-order condition is satisfied as well.

difference between essentially static and fully dynamic decision-making modes. By allowing firms to wait, it dramatically extends the investment opportunity set and modifies the net present value of investment. For example, empirical observations indicate that firms often find it in their interest to wait, even though the NPV of the project is positive (Summers 1987). This simple model can easily explain, at least qualitatively, the otherwise puzzling observation that firms seemingly use surprisingly high hurdle rates in their investment decisions.

The intuition behind this main result can be understood by analyzing the nature of the option that the firm possesses. When the future is uncertain, the option to wait is valuable, because over time firms learn new information and avoid investing when cash flow levels decline. By proceeding with its investment, the firm gives up this option to wait. Thus, the profit from the investment must be sufficiently attractive to trade off this option value. The optimal threshold X_I is such that the firm is indifferent between holding onto an option by waiting a bit longer and exercising its option now.

The comparative statics of the investment threshold, X_I, with respect to all the primitive parameters of the model, as we show in Table 2, are intuitive. The threshold increases with the investment cost I, inasmuch as a larger investment expense requires firms to become more certain that the state of the world is good. It also increases with cash flow volatility σ, because greater volatility increases the marginal value of waiting. That is, the distribution of future cash flows has fatter tails, and a delay can bring valuable information about changes in the probability of realizations of increasingly bad states of nature. The threshold also increases with respect to the discount rate r, because a higher discount rate lowers the present value of future cash flows relative to the cost of investment and increases the value of the waiting option. Finally, the threshold decreases with the growth rate μ, because a higher growth rate increases the value of potentially lost cash flows and thus makes waiting more costly.

A comparison of the optimal decisions in the dynamic and static (NPV) investment models, which we also include in Table 2, shows that even though both thresholds decrease with the growth rate, their difference increases: for higher growth opportunities firms wait longer compared to the standard NPV criterion. A comparison of values shows that at higher levels of the asset volatility, investment cost, and risk-free rate, the flexibility contribution becomes increasingly relatively valuable. In all these cases, the firm has more to lose by exercising its investment decision earlier.

The quantitative implications of this model are economically significant. For example, consider a project with the following reasonable characteristics: $X_0 = 1$, $I = 25$, $r = 0.05$, $\mu = 0.02$, and $\sigma = 0.25$. For these parameters, the take-it-or-leave-it NPV is positive. However, it is optimal for the firm to wait until the cash flow level increases from $X = 1$ to approximately $X_I = 2.4$ (a substantial increase of 140%). In this case, the value increases by 84%, from about 8.3 to 15.4. Another way to look at this example is to consider hurdle rates (or internal rates of return). In the NPV analysis, the value of the hurdle rate, r_H, that produces NPV equal to zero is 6%. If we take the option into account, the hurdle rate increases to 11.6%.

Figure 1 further shows that the hurdle rate in the presence of flexibility increases at a higher risk-free interest rate and asset volatility, and quantifies the impact of these two variables (using the same benchmark parameter we detailed previously). An uncertain environment in particular has a steep impact. For annual levels of volatility in the range of 0.40-0.60 (practically, reasonable for many applications), the hurdle rate is between 15% and 25% and can easily fit the ranges that have been reported empirically (e.g., Summers 1987). Thus, the model has a very good chance of explaining the stylized fact that firms wait too long (i.e., that implied hurdle rates are too high relative to the levels we would expect from applying the NPV analysis).

At this stage, it is useful to consider the specific economic ingredients that go into the model, because they are ubiquitous in a dynamic contingent claims paradigm. First, *flexibility* is critical: If the firm cannot wait (e.g., because competitive pressure makes the investment truly a take-it-or-leave-it-offer), the option value is immaterial. Second, the investment cost is *fixed*. Of course, what is important here is that the investment cost has a fixed component, as variable costs can be added easily. The presence of the fixed cost gives rise to an inaction region, in which firms prefer learning about the state of the world rather than making active investment decisions. Third, the investment is at least partially *irreversible*. If the firm can easily retract and salvage the present value of investment costs, then the real option is not valuable. In practice, all three components are present to some degree, and the framework enables us to study the extent to which a change in any of them changes the importance of the real option.

2.2.3 Discussion and Extensions

The simple models proposed by McDonald and Siegel (1986) and Brennan and Schwartz (1985) ushered corporate finance research into a new environment and in turn have been extended in various dimensions. Methodologically, this paradigm underlines the importance of the dynamic nature of both the environment and the decisions. In addition, because it is built on a dynamic setup and realistic option-pricing models, the framework provides the possibility of studying quantitative implications, such as how long firms actually wait to invest or the implied range of hurdle rates.

All the other examples we consider in this section are based on the fundamental idea that the option to wait before committing to a financial or investment policy is valuable. We turn our attention to the realm of investment and financing decisions extensively studied in corporate finance. Dixit and Pindyck (1994) consider various extensions of the basic real option model to other investment decisions and environments. A growing literature stream within corporate finance has been applying the real option framework to explore corporate investment decisions in depth. For example, Childs, Ott, and Triantis (1998) compare sequential and parallel investments in mutually exclusive projects. In line with the intuition we have developed, when projects' values are highly correlated, sequential investment is superior, because learning about the value of one project allows one to infer the value and thus the soundness of investment of another project.

Other important extensions that have only recently attracted attention of researchers include parameter uncertainty (Décamps, Mariotti, and Villeneuve 2005, 2009; Klein 2009), applications to real estate and leasing (Grenadier 1995), strategic competition (Grenadier 2002), agency cost and information asymmetry (Grenadier and Wang 2005), time-inconsistent preferences (Grenadier and Wang 2007), sequential investments and technology adoption (Grenadier and Weiss 1997), the dynamics of mergers and acquisitions in oligopolistic industries (Hackbarth and Miao 2012), asymmetric taxation and capital gains (Morellec and Schürhoff 2010), the impact of uncertainty on the probability of investment (Sarkar 2000), and the role of systemic risk and cost uncertainty (Sarkar 2003).

2.3 Optimal Static Capital Structure

The next application of the contingent claims paradigm we consider is the optimal financial structure of the firm. We start by building a formal trade-off capital structure model, then discuss the far-reaching economic implications of this setup.

2.3.1 Unlevered firm value

As in the case of a real options investment model, we assume that X is cash flow, generated this time from the assets in place owned by a firm, which also follows a geometric Brownian process:

$$dX_t = \mu X_t dt + \sigma X_t dW_t^{\mathbf{Q}}. \tag{2.7}$$

To concentrate on the firm's capital structure policy, we decouple investment and financial considerations by assuming that investment decisions are exogenous and investment and operating costs do not affect cash flows. The cash flow X in this setup is then equivalent to earnings before interest and taxes (EBIT). Because all the firm's future cash flows are generated by assets in place in perpetuity, the total asset value can be written as $\frac{X_t}{r-\mu}$.

The government levies a corporate tax on income at the constant proportional rate τ. If the firm's capital structure consists only of equity, then the value of equity at any time t, S_t, equals the value of the unlevered firm:

$$S_t = \frac{(1-\tau)X_t}{r-\mu}. \tag{2.8}$$

There are many realistic features that can be incorporated in the model at the cost of simplicity. For example, taxes are typically asymmetric in the real world, so that profits are taxed at a higher rate than losses, either explicitly or indirectly through carry-back and carry-forward provisions of the tax code. In addition, a personal tax is levied on distributions to claimholders. Although we ignore these considerations in our exposition, it is important to stress that some of them may have non-trivial quantitative implications.

Corporate debt provides an obvious way of lowering the tax bill, because interest payments are universally considered expenses by tax authorities and excluded from income taxation.[5] For simplicity, we consider perpetual debt that promises a constant coupon flow c. Assuming that the cash flow residual (i.e., free cash flow left after paying interest and taxes) is continuously paid out to shareholders, the payout flows to equityholders and debtholders are, correspondingly, $(1-\tau)(X_t-c)$ and c.[6]Thus, the total gain from tax benefits to debt is τc per unit of time.

2.3.2 Default

The cost of debt is that the firm may default on its debt payments later. The exact specification of a default model depends on institutional and legal framework. We consider two realistic scenarios for conditions in which default occurs and two scenarios describing potential firm outcomes upon default.

Optimal vs. Liquidity Default. Similar to the real options investment model, X is the only state variable, so default occurs only when X becomes sufficiently low and hits the default threshold, denoted X_D, for the first time from above. The determination of this threshold depends on the economic mechanism of default. Two contrasting scenarios are frequently considered in theoretical work. In the optimal or "endogenous" default case, equityholders (or managers representing them) choose a default threshold to maximize equity value. In other words, default is a real option that equityholders possess and by defaulting they exercise this option. As will soon become clear, in this case the effective assumption is that equityholders have "deep pockets", i.e. they have access

[5]The question of why coupon payments to debtholders are considered expenses, and thus receive a favorable tax treatment, whereas dividends to stockholders do not, is interesting in itself. This distinction seems likely to be the result of historical forces at the time the tax rules were being developed, rather than any deep economic reasoning pertaining to contemporary economic or business circumstances. The further discussion of this important issue is beyond the scope of this review.

[6]We thus assume away a proactive cash policy, whereby firms can retain earnings in anticipation of future needs. Incorporating a cash policy in this setup reduces tractability, though studies of cash policy in the presence of debt in different setups suggest that it is an important issue (see e.g., Acharya, Davydenko, and Strebulaev 2012; Anderson and Carverhill 2012).

to external funds to cover coupon payments if needed. The solution method to the problem of finding the optimal default is exactly the same as for the investment problem of Section 2.2. In the liquidity or "exogenous" default case, the firm defaults either because it violates net worth covenants or because the firm and equityholders have no spare cash to cover their current interest payments.

Both scenarios and their various amalgamations are realistic. For example, even though firms can raise additional equity funding to pay debtholders, it may be costly or difficult due to timing constraints or covenants in the debt contract. With some foresight, these elements can be modeled parsimoniously within one model as $X_D = \gamma c$. In the benchmark liquidity case, γ is equal to 1, implying that the firm cannot raise any new external funds, and instead defaults the moment its cash flow is lower than its promised coupon. In the optimal default case, γ is a control variable determined by maximizing shareholders. If the optimal γ is lower than 1, the "deep-pockets" assumption becomes clear. Shareholders have the means to pay debtholders, as long as they want to keep their option alive. Alternatively, this assumption can be replaced by the realistic feature that it is finitely costly to raise additional funds in distress.

Liquidation vs. Reorganization in Default. The economic cost of debt is modeled as the partial loss of future cash flows in default. We assume that this cost affects all future cash flows proportionally at the rate α.

In default, the firm is liquidated or reorganized, and the proceeds are distributed to the claimants according to the absolute priority rule (APR). In the simple capital structure case with only two claimants, debt and equity, debt has seniority over equity, meaning that until it is paid in full, equityholders are not compensated. For the endogenous default scenarios we consider, the APR in fact implies that equityholders recover no value in default. Empirically, deviations from the APR are frequently observed in the U.S., such that junior claimants recover non-zero value in bankruptcy before more senior claimants are satisfied in full (e.g., Franks and Torous 1989). Default outcomes also depend in large part on the institutional and legal framework. Such deviations can be easily accommodated as an exogenous model feature (assuming a specific liquidation split) or as an institutionalized bargaining game in default.

Two alternative scenarios can occur in default. In the case of liquidation, assets are redeployed elsewhere, and debtholders receive the present value of the after-tax cash flow stream. Our treatment of liquidation implies that debtholders partially recover the value of assets effectively as new equityholders, but they do not lever up the assets again by issuing new debt. This case closely corresponds to Chapter 7 of the U.S. Bankruptcy Code or the administrative receivership procedure in the U.K. Therefore, the value of the firm at default, V_D, can be written as:

$$V_D(X_D) = (1-\alpha)(1-\tau)\frac{X_D}{r-\mu}. \tag{2.9}$$

Alternatively, default can lead to reorganization rather than liquidation, so that new equityholders establish a new firm with the remaining assets in place, and issue debt to maximize its value. The value in default then depends on the resulting value of the new firm. Such a reorganization is similar to bankruptcy outcomes under the Chapter 11 of the U.S. Bankruptcy Code, and is considered in Section 2.3.6. For now, we assume that the firm is liquidated. The critical difference between these two cases is whether the tax benefits of debt get lost in default. If the firm is liquidated, but the assets are sold to another party that intends to follow optimal capital structure and pays fair value for it, the resulting value in default should correspond to the value in the case of reorganization. Practically, the tax shield gets lost only for the duration of the Chapter 11 bankruptcy process. Thus, liquidation and reorganization can be viewed as generating low and upper boundaries for the firm value in default, respectively. Broadie, Chernov, and Sundaresan (2007) consider an interesting extension that distinguishes between liquidation (Chapter 7) and reorganization (Chapter 11). In their model, the firm that files for bankruptcy is granted an automatic stay (no debt payments are disbursed, although they accumulate). If the company's condition improves, the firm reorganizes and partially pays all the accumulated debt obligations. If the condition deteriorates, the firm is transferred into liquidation. One realistic implication of their framework is that although the probability of default and bankruptcy is higher, the probability of eventual liquidation is lower.

2.3.3 Equity and Debt Valuation

As in the real options investment model, it is helpful to consider an Arrow-Debreu contingent claim that pays \$1 only when the level of cash flow reaches the default threshold for the first time from above. The value of the Arrow-Debreu claim at any time prior to the time of default can be shown to be:[7]

$$A(X_t, X_D) = \left(\frac{X_t}{X_D}\right)^{\xi_1}, \tag{2.10}$$

where ξ_1 is the negative root of the fundamental quadratic equation:

$$\xi_1 = -\frac{1}{\sigma^2}\left(\mu - \frac{\sigma^2}{2} + \sqrt{\left(\mu - \frac{\sigma^2}{2}\right)^2 + 2\sigma^2 r}\right). \tag{2.11}$$

At any date prior to default, the value of levered equity can be written as:

$$S(X_t) = \left(\frac{X_t}{r-\mu} - \frac{c}{r}\right)(1-\tau) + A(X_t, X_D)\left(-\frac{X_D}{r-\mu} + \frac{c}{r}\right)(1-\tau). \tag{2.12}$$

The first term in Equation (2.12) is the after-tax levered equity value in perpetuity, assuming no default. The second term takes default into account. By defaulting, shareholders forgo future cash flows in exchange for discontinuing interest payments.[8] To find the optimal level of X_D, we apply the smooth-pasting condition:[9]

$$\left.\frac{\partial S(X_t)}{\partial X_t}\right|_{X_t = X_D} = 0. \tag{2.13}$$

[7]From now on, we assume that t is a time point prior to default: $t < \min\{s : X_s \leq X_D\}$. Note that the values of basic contingent claims in this and the real options models are different, because in the investment model the payoff occurs only when the upper trigger is reached, and here it takes place only when the lower trigger is reached.

[8]In the exogenous default case, equityholders are also entitled to the residual asset value after debtholders have been satisfied in full. The residual is not zero for very small values of default costs. Although this slightly changes the formulation of the problem, we ignore this scenario here for simplicity.

[9]It is easy to confirm that, in this simple problem, the smooth-pasting condition is equivalent to the maximization of shareholder value with respect to X_D. Stokey (2009) provides a formal treatment of smooth-pasting in a more general setup.

The optimal default boundary is then:

$$X_D = \frac{\xi_1}{\xi_1 - 1} \frac{r - \mu}{r} c. \tag{2.14}$$

Thus, the default boundary is proportional to the coupon, so we can write $X_D = \gamma_D c$, where $\gamma_D = \frac{\xi_1}{\xi_1 - 1} \frac{r-\mu}{r}$. Because $\xi_1 < 0$, it is easy to show that $\gamma_D < 1$, and therefore in the endogenous default case shareholders find it optimal to keep their default option alive longer than they would in the benchmark liquidity default case, where $\gamma = 1$. For a value of γ below one, keeping the default option alive requires external funds to pay the coupon in distress (defined as $X < c$), which provides the rationale for the rights issues or, more generally, equityholders' "deep pockets." In practice, it is likely that γ is the weighted average of the values in these two opposite cases. A reasonable scenario is a partial liquidity case, where equityholders incur issuance costs to access external funds in distress.

An important observation is that the optimal default boundary X_D is independent of the default cost α. In making a default decision, equityholders do not internalize debtholders' value, and only debtholders, as residual equity claimants, bear any default costs. As we emphasize soon hereafter, this ex ante vs. ex post conflict of interest plays an important role in initial capital structure decisions.

The value of perpetual debt promising coupon c, $D(X_t)$, is (recall that we assume the case of liquidation in the case of default):

$$D(X_t) = \frac{c}{r} - A(X_t, X_D)\frac{c}{r} + A(X_t, X_D)(1 - \alpha)(1 - \tau)\frac{X_D}{r - \mu}. \tag{2.15}$$

The first term is the perpetuity value of risk-free debt. The second term is the present value of interest payments that debtholders lose in the case of default. The final term is the present value of assets that debtholders recover in liquidation.

2.3.4 Optimal Capital Structure

The shareholders of an all-equity firm maximize their equity value—that is, in this case, firm value—at the initial date 0 by choosing an optimal capital structure. We assume in this static capital structure environment that the decision to issue debt is made only once. At any subsequent date, equityholders may default but are not allowed to refinance. Section 2.4 considers the first-order extension that allows for refinancing.

Note that at date 0, by maximizing the total value of future equity and debt, shareholders internalize the value of future debtholders' claims. This finding is not inconsistent with the result on shareholders' default decision. The crucial distinction between debt at date 0 and debt after it has been issued centers on *future* versus *outstanding* claims. In the latter case, debt issuance is a sunk decision (and the amount contributed by debtholders is a sunk benefit) that is not taken into account. In the former case, debt is still to be issued and priced by investors, so equityholders take the debt claim into account by maximizing total firm value. Rational debt markets reflect the future conflict of interest in pricing the firm's claims. Intuitively, it would be better for equityholders to commit ex ante to maximizing firm value rather than equity value at any time, not only at issuance. We realistically assume that such a commitment is not contractible, though in practice various covenants in debt indentures attempt to substitute for such a commitment.

Shareholders maximize total firm value, $V(X_0)$, subject to the constraint $X_D = \gamma c$. There are two approaches to writing firm value, V, after debt has been issued. First, it can be written as the sum of debt and equity, $V = D + S$, using Equations (2.12) and (2.15). Second, it can be written as the sum of all the components that make firm value deviate from the after-tax unlevered asset value: $V = F + TB - DC$, where F is the unlevered firm value (equivalent to equity value when $c = 0$), TB is the present value of the tax benefits of debt, and DC is the present value of default costs. It is instructive to write the firm value in both ways. In the second approach, the component values are:

$$
\begin{aligned}
F(X_t) &= (1-\tau)\frac{X_t}{r-\mu}; & (2.16)\\
DC(X_t) &= A(X_t, X_D)\alpha(1-\tau)\frac{X_D}{r-\mu}; & (2.17)\\
TB(X_t) &= \tau\frac{c}{r}(1 - A(X_t, X_D)). & (2.18)
\end{aligned}
$$

To find the optimal coupon level, shareholders maximize the date-0 firm value, $V(X_0)$, with respect to the coupon c, subject to the default condition $X_D = \gamma c$. From the first-order condition, the solution to this problem is:

$$
c^* = \frac{1}{\gamma}\left[\frac{1}{1-\xi_1}\frac{\tau}{(1-\tau)\alpha\gamma\frac{r}{r-\mu} + \tau}\right]^{-\frac{1}{\xi_1}} X_0. \tag{2.19}
$$

In the endogenous default boundary case, we can find the optimal coupon by substituting for the value of γ from Equation (2.14):

$$
c^*(\gamma_D) = \frac{r}{r-\mu}\frac{\xi_1 - 1}{\xi_1}\left[\frac{1}{1-\xi_1}\frac{\tau}{\frac{\xi_1}{\xi_1 - 1}(1-\tau)\alpha + \tau}\right]^{-\frac{1}{\xi_1}} X_0. \tag{2.20}
$$

Thus, the optimal coupon is proportional to the cash flow level, implying that size of the firm is irrelevant in this specification.

2.3.5 Comparative statics

In Table 3 and Figures 2–6, we present comparative statics with respect to the primitive parameters of the model. To understand the economics that underlie these comparative statics, we need to emphasize the main economic mechanism at play: Changes in the primitive parameters shift the term structure of the expected cash flows stemming from the tax benefit and default cost, because these cash flow streams occur at different times. The table and figures also allow us to compare endogenous and exogenous default cases. To understand the intuition behind the differences between these two cases, note that while in the endogenous case the equityholders exert control by choosing two parameters (the coupon level at the outset and the default boundary in the future), in the exogenous case the choice of coupon dictates the choice of the default trigger. For each coupon level, the distance to default is smaller and thus the present value of the default costs is larger in the exogenous case. This distinction accounts for the differences in their comparative statics, as well as for the consistently lower values of optimal coupon and default boundary in the exogenous default case.

A challenging empirical problem for interpreting the coupon level as a main proxy for financial policy is that firms of different size are not directly comparable. Instead, both empirically and theoretically, it is more common to use financial leverage. At any point in time, we can define the market leverage ratio as:

$$L(X_t) = \frac{D(X_t)}{V(X_t)}. \tag{2.21}$$

In practice, the market values of debt often are unavailable, and the quasi-market leverage ratio is used instead:

$$QML(X_t) = \frac{D(X_0)}{D(X_0) + S(X_t)}, \tag{2.22}$$

where $D(X_0)$ indicates the original (book) value of debt. When the firm is not in distress, the two ratios are typically very close in value. If debt is issued at par, then the date-0 leverage and quasi-market leverage ratios are equal. For the purpose of this section, we consider the date-0 leverage only.

We are also interested in measuring the effect of the decision to lever up on the value of the firm. To do so, we consider the relative value differential between levered and unlevered firms at date 0, $\frac{V(X_0)-F(X_0)}{F(X_0)}$.

Consider a variation in the growth rate parameter. A higher growth rate increases expected future cash flows and intuitively increases the firm's desire to take on more debt to increase its tax benefits. In the endogenous case, the firm responds by implementing exactly this policy, while simultaneously lowering the default threshold per unit of coupon (as Table 3 shows, γ_D is decreasing in μ). In the exogenous case, however, the firm does not explicitly control the default trigger, so a higher coupon implies a greater probability of default early on, which may lead the firm to lose out on the bright (but more distant) future. The firms makes an optimal decision by balancing this trade-off. The effect of a higher growth rate may thus actually lower the optimal coupon. For the same reason, leverage and the value differential increase (decrease) with the growth rate in the endogenous (exogenous) case.

In the endogenous default case, the leverage ratio increases with the interest rate, even though the optimal coupon is lower. To understand the difference, note that the behavior of the leverage ratio depends on the relative sensitivities of equity and debt values to the variation in the underlying parameters. As the interest rate increases, all future cash flows are discounted at a higher rate, but equity is affected more, because most of its value comes from future cash flows. In other words, it can be said that equity has a greater duration than debt and thus it is more sensitive to the interest rate variation. Even though the coupon decreases (because tax benefits, expected in the more distant future, are more affected), the impact of the lower debt value is negated by the even lower value of equity. The effect is obviously more pronounced in the exogenous default case, because both forces act in the same direction.

Asset volatility exerts two effects on optimal financial policy. On the one hand, higher volatility increases the risk of default, which implies a lower optimal coupon. On the other hand, higher volatility increases the likelihood of the firm finding itself in very good states of the world, where potential tax benefits are high. As Figure 4 and Table 3 show, in the endogenous case the first effect dominates for low values of asset volatility and leads to a smaller coupon, while the second effect dominates for larger values. In the exogenous case, the risk of default naturally plays a more prominent role. The leverage ratio is a decreasing function of asset volatility, because higher asset volatility makes equity, as a call option, more valuable, but debt, as a short position in a put option, less valuable. Finally, as expected, the variables of interest decline with higher default costs and increase with a higher tax rate.

2.3.6 Optimal Capital Structure with Reorganization

We can derive another empirically relevant case when debtholders reorganize a defaulted firm and lever it up again. The valuation of equity is unchanged and given by Equation (2.12), because equityholders do not internalize the actions of debtholders upon default after a debt issuance. Therefore, the default boundary is also the same, $X_D = \gamma_D c$, and γ takes the same values we previously established for the endogenous and exogenous cases.

The value of perpetual debt, $D(X_t)$, is thus

$$D(X_t) = \frac{c}{r} - A(X_t, X_D)\frac{c}{r} + A(X_t, X_D)(1-\alpha)\left(S(X_D) + D(X_D)\right), \tag{2.23}$$

where $V_D(X_D) = S(X_D) + D(X_D)$ is the value of the new firm. The first-order homogeneity of the claims means that we can write debt value as:

$$D(X_t) = \frac{c}{r} - A(X_t, X_D)\frac{c}{r} + A(X_t, X_D)(1-\alpha)\frac{X_D}{X_0}\left(S(X_0) + D(X_0)\right), \tag{2.24}$$

and therefore, the debt value at date 0 is:

$$D(X_0) = \frac{\frac{c}{r} - A(X_t, X_D)\frac{c}{r} + A(X_t, X_D)(1-\alpha)\frac{X_D}{X_0}S(X_0)}{1 - A(X_t, X_D)(1-\alpha)\frac{X_D}{X_0}}. \tag{2.25}$$

Equityholders still maximize the total value of the firm at date 0 by choosing the coupon level, subject to the appropriate default boundary condition. Compared with the liquidation case, the comparative statics are qualitatively similar, but a reorganization option effectively reduces the cost of default and leads to a higher reliance on debt. For the same reason, the difference between endogenous and exogenous default cases is smaller. We compare these two cases quantitatively in Section 2.3.8.

2.3.7 Discussion

The trade-off capital structure model is helpful for illustrating the comparison between traditional static and contingent claims paradigms. For clarity, it is worthwhile to briefly review the historical background of capital structure studies. Following the landmark indeterminacy result of Modigliani and Miller (1958) and its extension to cover tax benefits in Modigliani and Miller (1963), the debate on optimal corporate capital structure centered on what has since become known as static trade-off theory. The general view in the 1960s and the 1970s was that static trade-off theory offered a first-order explanation of the financial structure of firms. Miller (1977) criticized this view, arguing that bankruptcy costs, an obvious culprit and counterweight to offset the tax benefits of debt, are likely too small to explain why firms appear insufficiently levered. In Miller's words, tax benefits versus bankruptcy costs are like "a horse and rabbit stew." Miller (1977) suggests personal taxes as one resolution. DeAngelo and Masulis (1980) provide another trade-off-based explanation that relies on alternative tax shields, such as leasing and depreciation. The modified consensus that emerged as a result of these efforts came to the conclusion that optimal capital structure is the result of the trade-off between the tax advantages of debt and various leverage-related costs.[10]

No formal model is required to understand the economic forces in this trade-off model. A higher corporate tax rate should lead to higher optimal leverage, whereas a higher personal income tax should work in the opposite direction. Higher profitability increases taxable cash flow and thus increases corporate demand for debt, whereas higher bankruptcy costs reduce the debt advantage. Even though such a trade-off naturally yields an interior optimal capital structure, a static model has difficulty quantifying relevant economic forces. At the least, the timing of cash flows stemming from the tax benefits of debt and the timing of cash flows lost due to bankruptcy costs do not coincide. A static model is therefore likely doomed in its attempt to quantify this cost-benefit trade-off, because the present values of these asynchronous costs and benefits are difficult to compare.

Therefore, the "horse and rabbit stew" conundrum and subsequent ad hoc attempts to address it have been based largely on intuitively reasonable, qualitative observations rather than on careful quantitative analysis. Such qualitative arguments cannot gauge the size of the gap between theory and reality, nor can they help researchers understand the quantitative effects of relevant forces, such as asset market volatility or interest rate changes, even though these issues have been widely recognized as important.

In light of these developments, the contribution of the contingent claims model becomes clear. At face value, it does not propose new economic mechanisms, because the basic economic forces are identical to those in the traditional trade-off model of tax benefits and bankruptcy costs. Although

[10]It is important to note that Miller (1977) and DeAngelo and Masulis (1980) study adjustment of capital market prices as a result of the trade-off between net tax benefits of debt and debt-related costs.

Bradley, Jarrell, and Kim (1984) discuss dynamic intuition in a static model, the honor of being the first in this area likely belongs to Kane, Marcus, and McDonald (1984), who apply option pricing methods to study the extent of the tax advantage of debt. Fischer, Heinkel, and Zechner (1989) then built a fully-fledged dynamic model that allows for dynamic capital structure. Yet it was a contribution by Leland (1994) that really started the contingent claims revolution in corporate finance. The model that Leland (1994) developed is an amalgam of all the preceding contributions. Its greater importance, in terms of the future development of the field, was in its simple and yet clear formal treatment of trade-off theory. It developed a relatively robust contingent claims model of capital structure, based on the intuition of the static model, and then quantified optimal leverage ratios that firms should choose if they live in the dynamic world described by the trade-off model. Leland's model further exploited option pricing models to study corporate credit risk and thus was fundamental for the subsequent development of contingent claims credit risk models as well.

In contrast to prior models, Leland's is simple enough to be easily comparable to static trade-off models. It is thus well-suited to illustrate the power of contingent claims analysis. This framework confers the basic advantage that the Arrow-Debreu-type security $A(X, X_D)$ can be specified using a well-established asset pricing paradigm. The value of A depends not on ad hoc assumptions about the costs and benefits of debt in various two-period static models,[11] but rather on primitive variables (e.g., asset volatility, riskless interest rates) that can be estimated or approximated reasonably well. Consequently, costs, benefits, and primitives determine the value of tax benefits and bankruptcy costs, the resulting value of the firm, and optimal financial decisions. Contingent claims analysis enables us to investigate the comparative statics, both qualitatively and quantitatively, that relate financial structure control variables to the primitive economic parameters. In turn, we can quantify their effect on capital structure and judge whether reasonable capital structures can be obtained for any realistic range of parameters. Using this formulation, it is also easy to understand what other factors might be missing and how their effects can be incorporated. If the calibration shows that the model performs poorly, not only can the model be rejected, but it is also possible to explore

[11]Examples would include quadratic costs and linear benefits, which are used entirely for tractability and to obtain an interior solution to the optimal capital structure problem, rather than for any reasons pertaining to realism.

whether the modeling of the security A needs to be modified and additional economic factors should be considered.

What is important is that, by modeling the security A in this fashion, we can facilitate modeling of various costs and benefits time-consistent. Cash flows arising from tax benefits and bankruptcy costs, as well as other leverage-related costs and benefits, do not occur at the same time. Although bankruptcy costs potentially are substantial in the event of an eventual default, the expected present value of bankruptcy costs is much smaller, so they need to be adjusted for the time-weighted probability of default. Thus, A parsimoniously reflects the term structure of the default distribution over time, which is an important consideration that separates contingent claims models from their simple static counterparts.

As we show in Section 2.3.8, for the endogenous default scenarios, and to a lesser extent for the exogenous case, the model confirms the well-accepted "horse and rabbit stew" conundrum, in which optimal leverage, as implied by reasonable parameters, is too high to explain actual firm behavior in the cross-section. For example, in the endogenous default scenario, the model produces a leverage ratio in excess of 70%, as opposed to the average leverage ratio of 25% that we observe in reality.

These findings, first clearly demonstrated by Leland (1994), are as important to the development of capital structure, and perhaps to the whole field of corporate finance, as the description of the equity premium offered by Mehra and Prescott (1985). Leland not only showed that the dynamic version of the static-trade-off model could not replicate observed leverage, but he laid the groundwork for understanding the sources of this failure by constructing a simple model with well-defined assumptions. Naturally, most subsequent research based on this and earlier contingent claims models has explored the role of various assumptions. We highlight some of the more important ones in subsequent sections.

2.3.8 Quantitative implications

In this section, we explore the quantitative implications of the capital structure we have developed thus far. We do this in a couple of ways. First, we start by considering a particular set of parameters, for which we choose reasonable values relying on existing calibration and empirical studies. Second, we quantify the comparative statics of the leverage ratio with respect to the variation in the empirically relevant parameters. We concentrate on the leverage ratio at date 0, consistent with most of the extant theoretical literature.[12]

The benchmark set of parameters is as follows: interest rate, r, is 0.05, the growth rate, μ, is 0.02, asset volatility, σ, is 0.25, the corporate tax rate, τ, is 0.2, and the default cost parameter, α, is 0.15. For many of these parameters, the direct empirical estimation is clearly difficult; in addition, there is substantial cross-sectional heterogeneity. Some empirical evidence enables us to anchor the benchmark values, though. For example, Schaefer and Strebulaev (2008) estimate the average asset volatility of firms that issued bonds to be 0.23, and default costs are approximately 10–20% of the total firm value at default, according to Andrade and Kaplan (1998), or 15–30%, according to Davydenko, Strebulaev, and Zhao (2012). Although the U.S. federal corporate tax rate is 0.35, the effective tax rate is likely smaller, reduced by, for example, personal tax rates (Graham 2000).

As Table 4 shows, for this benchmark set of parameters, the leverage ratio in the endogenous default case varies between 70% (liquidation) and 80% (reorganization) – far in excess of the observed leverage ratios. For example, the equally-weighted quasi-market leverage ratio in the COMPUSTAT universe consistently appears within the 20–25% range. The leverage ratios in the exogenous default case are substantially smaller, at 25–28%. However, it is unlikely that the model can explain the actual observed capital structure data, because in practice, firms have substantial flexibility to raise additional funds (or use retained earnings) to pay down their interest. Firms do not realistically default every time their cash flow drops below the promised interest payment.

[12] Needless to say, in the model with the leverage decision only at date 0, the leverage ratio attenuates over time, conditional on no default. In Section 2.5, we discuss the behavior of the leverage ratio over time in a dynamic model of capital structure decisions.

At the same time, this model can easily account for differences between large and small firms. Large firms are likely more flexible in their outside sources of financing, and temporary liquidity problems should have less significance for them. For smaller firms, cash flows that are too low to cover interest can be a much more troublesome sign. In turn, large firms likely have more debt capacity.

Table 4 also shows that the value differential between levered and unlevered firms can be substantial. In the endogenous case, the firm's value increases by 11–13%, though the impact is much more modest in the exogenous case.

The rest of the table illustrates how leverage changes when we vary the main mode parameters. In the endogenous case, the leverage ratio is substantially in excess of empirically observed ratios, even for extreme parameter values. For example, leverage is 48–50% when half of the firm is lost in default. Note that leverage is still high, even if the tax rate is low: when the effective tax rate is just 10%, the leverage ratio is still 59–65%. These results contrast with the substantially lower leverage ratios in the exogenous case. Empirically, if it is possible to identify exogenous variation in the values of γ, the model provides clear testable predictions. Because the endogenous case is likely to be more realistic for a large cross-section of firms, these results clearly suggest that the model produces much higher leverage values than we observe. In the next several sections, we consider several extensions of the basic model that lead to a considerably lower optimal leverage at date 0.

2.4 Optimal Dynamic Capital Structure

The assumption that firms can raise debt financing only once is unrealistic. In theory, a static debt policy implies that as firms grow larger over time, their leverage attenuates, leading to an unreasonably negative relationship between leverage and firm size. In practice, firms' financial policies are dynamic, and leverage adjusts in response to the changing environment. In this section, we formalize this important idea by building a benchmark model of optimal dynamic capital structure.

2.4.1 Model Setup

The intuition for this model is based on two fundamental properties of the dynamic environment. First, firms should take into account any expected changes to their future cash flows and opportunity sets when forming their current financial policy. The opportunity to adjust their leverage at a future date is likely to change firms' financial decisions today. Second, firms face various adjustment costs (e.g., the transaction costs of raising external financing), and even though these costs may be small relative to the total amount raised, their effect on optimal policies can be substantial. The impact of adjustment costs on optimal decisions has been studied extensively in macroeconomics, such as in the context of menu cost models, in which firms do not adjust prices continuously because of the small transaction costs. The generic outcome of those models is that there is a region of a state variable (e.g., demand for a firm's product), called an inaction region, in which the firm prefers *optimally* not to make active changes in its effort to reduce the expenses associated with any adjustment. Mankiw (1985) shows that even very small menu costs can have a profound impact on aggregate macroeconomic variables. One of the implications of dynamic corporate finance is that a similar intuition fully applies to corporate decisions and that the magnitude of firms' response is similarly large.

To introduce dynamic capital structure decisions to the model in Section 2.3, assume that in addition to a default option that firms find valuable to exercise in bad states of the world, they also possess a refinancing option that they may find valuable to exercise in good states.[13] As the cash flow level, X, grows, the present value of the default costs decreases (because firms are further away from the default threshold). A higher cash flow level also implies a more punitive effective tax rate (because fewer profits are shielded from the corporate tax by a constant coupon). The refinancing option is valuable, because it enables firms to restructure their capital structure at some upper threshold by raising more debt and restoring the balance between the costs of distress and the tax benefits of debt.

[13] Refinancing in a bad state of the world might be considered as well but is less important (default, of course, is an extreme version of refinancing). Realistically, firms in distress may need to find additional resources at an extra cost; we do not implement this potentially fruitful extension here.

Because raising new debt and refinancing existing debt are costly, the next logical step is to introduce the cost of refinancing. In the absence of any refinancing costs, firms adjust their outstanding amount of debt continuously in good states. We assume that this cost is proportional to the total amount of debt raised at any refinancing date. This assumption allows us to model the fixed cost of refinancing in a tractable manner because (as we show below) it preserves the first-order homogeneity property of the problem, also known as scaling. Under this cost structure, if the firm raised debt originally at X_0, it will optimally default when X reaches the low value of X_D, and it will refinance when X reaches the upper value of X_R, where $X_D < X_0 < X_R$.

Using the by now familiar strategy, we start by considering the appropriate Arrow-Debreu-type (hereafter, AD) contingent claims. In the static investment problem of Section 2.2, the appropriate AD contingent claim pays \$1 when the cash flow level reaches the upper threshold; in the capital structure problem of Section 2.3, the AD contingent claim is the mirror image, such that it pays \$1 when the cash flow level reaches the lower threshold. In the dynamic problem with two thresholds, we therefore introduce two AD contingent claims. The first, denoted $A(X, X_D | T_{X_D} < T_{X_R})$, pays \$1 when the cash flow level reaches the lower threshold X_D for the first time from above, if and only if this event happens *before* it reaches the upper threshold X_R (T_X indicates the first time it reaches value X). Similarly, $A(X, X_R | T_{X_R} < T_{X_D})$ denotes the value of a contingent claim that pays \$1 when the cash flow level reaches the upper threshold X_R for the first time from below, if and only if this event happens *before* it reaches the lower threshold X_D.

The market value of debt, raised at date 0, can be written at any date t (before either default or refinancing occurs) as:

$$\begin{aligned} D(X_t) &= \frac{c}{r} + A(X_t, X_D | T_{X_D} < T_{X_R}) \left(-\frac{c}{r} + (1-\alpha) V_D(X_D) \right) \\ &\quad + A(X_t, X_R | T_{X_R} < T_{X_D}) R(X_R), \;\; T_{X_t} \leq \min\left(T_{X_D}, T_{X_R}\right). \end{aligned} \tag{2.26}$$

The second term is a familiar payoff to debtholders in the event of default, where V_D depends on the outcome of default: liquidation (Equation (2.9)) or reorganization (Equation (2.23)). The last term is the debt payoff at refinancing, $R(X_R)$. There are two alternate assumptions for modeling $R(X_R)$. In one case, existing debt is recalled (at par, or its original book value at date 0, or at a premium) and new debt is issued in its place. Taken literally, the firm issues callable debt with an indenture agreement that prevents it from issuing any further debt. In another case, the old debtholders' claim is diluted by the issuance of new debt with a *pari passu* provision, such that all outstanding debt issues have equal seniority. Although the exact specifications strongly reflect the desire for tractability, both of these assumptions are realistic; in addition, many variations can be considered (likely without significantly affecting any of the model's economic results). Consider the callability specification, such that $R(X_R) = -\frac{c}{r} + (1+\omega)D(X_0)$, where ω is a premium the firm must pay for the right to recall its debt. Then, from Equation (2.26), and assuming the firm is liquidated in default, we can rewrite the value of debt at date 0 as:

$$D(X_0) = \frac{\frac{c}{r} + A\left(X_0, X_D | T_{X_D} < T_{X_R}\right)\left(-\frac{c}{r} + (1-\alpha)(1-\tau)\frac{X_D}{r-\mu}\right) + A\left(X_0, X_R | T_{X_R} < T_{X_D}\right)\left(-\frac{c}{r}\right)}{1 - A\left(X_0, X_R | T_{X_R} < T_{X_D}\right)(1+\omega)}. \tag{2.27}$$

The value of equity before either default or refinancing occurs is (not taking into account date-0 transactions):

$$\begin{aligned} s(X_t) \;=\; & \left(\frac{X_t}{r-\mu} - \frac{c}{r} + A\left(X_t, X_D | T_{X_D} < T_{X_R}\right)\left(-\frac{X_D}{r-\mu} + \frac{c}{r}\right)\right)(1-\tau) \\ & + A\left(X_t, X_R | T_{X_R} < T_{X_D}\right)\left(-\frac{X_R}{r-\mu} + \frac{c}{r}\right)(1-\tau) \\ & + A\left(X_t, X_R | T_{X_R} < T_{X_D}\right) S(X_R), \;\; T_{X_t} \le \min\left(T_{X_D}, T_{X_R}\right). \end{aligned} \tag{2.28}$$

The last two terms adjust for the refinancing point. Intuitively, from the equityholders' point of view, refinancing reflects two sequential transactions: The equityholders forgo future cash flows in exchange for coupon payments (the second line), then receive the total residual value of equity (the third line). Note the difference between $s(X)$ and $S(X)$: whereas $S(X)$ is the *total* value of equity, $s(X)$ takes into account the impact of transactions at the beginning of a refinancing cycle (date 0 in this case). At all values of X (apart from X_0, X_R, and other refinancing levels), $s(X) = S(X)$. At refinancing, equityholders pay off existing debtholders and issue new debt, so that the total value of equity is:

$$S(X_R) = -(1+\omega)D(X_0) + (1-q)D(X_R) + s(X_R), \tag{2.29}$$

where q is the proportional cost of debt issuance. The proportionality of the cost (as well as the nature of the cash flow stochastic process) assures the first-order homogeneity property. In other words, all the securities' values at refinancing are proportionately larger than the corresponding values at date 0. We thus can write the value of equity at T_R as:

$$S(X_R) = -(1+\omega)D(X_0) + (1-q)\frac{X_R}{X_0}D(X_0) + \frac{X_R}{X_0}s(X_0). \tag{2.30}$$

Equityholders maximize the total value of equity at date 0 (before debt is issued):

$$S(X_0) = (1-q)D(X_0) + s(X_0). \tag{2.31}$$

Their controls are c, X_D, and X_R. A subtle point about this optimization problem is that though equityholders cannot credibly commit to default at a particular level at date 0 in the endogenous default case (and thus X_D is still found from the smooth-pasting condition), by issuing callable debt they likely can commit effectively to recall at a certain cash flow level.

Commitment is valuable, so X_R can be obtained from the date-0 joint maximization of the total equity value with respect to c and X_R, rather than from the smooth-pasting condition. However, in the absence of a credible commitment device, X_R is obtained from the smooth-pasting condition (or rather, X_R and X_D can be determined from the joint smooth-pasting conditions). Here, we assume the former case, but the latter scenario leads to qualitatively similar results.

Another important point worth emphasizing is that unlike the problem of Section 2.3, default costs here enter the default decision of equityholders: Although equityholders do not internalize the value of outstanding debt, they internalize the value of all future debt issues, which of course depend on default costs.

2.4.2 Comparative Statics

Most of the economic forces in the fully dynamic model have not changed, so the qualitative predictions overall remain similar to those developed in Section 2.3.8. Figure 7 compares the behavior of the endogenous and exogenous default cases as debt issuance costs vary (we use the same benchmark set of parameters as in Section 2.3.8). For the endogenous case, the dependence of the optimal coupon and leverage on issuance costs is non-monotonic. A simple explanation of this seemingly counterintuitive behavior is that as adjustment costs increase, firms delay refinancing (as also illustrated), but conditional on refinancing, they issue more debt. This channel works only because the firm has the refinancing boundary as an additional control. As costs increase, the overall attractiveness of debt attenuates. In general, an extra degree of freedom implies that firms constantly face the trade-off of either issuing more at each refinancing point, while delaying refinancing, or else issuing less but refinancing more often. Interestingly, the refinancing boundary, X_R, is typically uniformly higher in the exogenous default case: Firms respond to a higher default probability not only by choosing a lower coupon but also by delaying refinancing in good states of the world.

Figure 8 shows the difference between dynamic and static leverage models with endogenous default for different levels of the debt issuance cost. To make these models comparable, we consider a static leverage model that is slightly different from the one in Section 2.3: The debt issuance cost is incurred at date 0, so the firm's maximization problem is identical in both cases. The first-order implication is that when firms acquire an additional degree of freedom, they lower their coupon levels and exhibit lower leverage ratios compared with the static version. This behavior is entirely intuitive, in that only reason firms pursue an aggressive capital structure policy at date 0 in the static case is their inability to issue more debt when cash flow levels increase. As an opportunity to recapitalize becomes available, they issue less debt initially to reduce the likelihood of default. For the same reason, equityholders opt to default later in the dynamic case, because their upside option is enhanced by the opportunity to refinance later.

Figure 8 also shows that adding associated with flexibility of adjusting leverage dynamically increases the value of the levered firm. This result is very general. The difference in values between dynamic and static capital structure firms increases in the growth rate μ, asset volatility σ, and tax rate τ, but it decreases with the default cost parameter, α, and issuance cost parameter, q. These relationships follow directly from the economic forces we have already discussed. For example, as the tax rate increases, future tax benefits become relatively more important, and dynamic flexibility adds more value.

2.4.3 Quantitative Implications

Table 5 reports the quantitative implications of the dynamic capital structure model for the two default cases we have considered. We assume that the call premium, ω, is zero. For comparison, we also provide predictions pertaining to the static capital structure model. To make these models comparable, we again assume that in the static model the firm pays the debt issuance cost at date 0.[14] The introduction of dynamic flexibility lowers the date-0 leverage ratio substantially. In the benchmark endogenous default case, the difference is 17%, though it ranges between 10% and 23% for the parameter constellations we consider. An important implication is that, although the leverage ratio is substantially lower, it is still significantly higher than the empirically observed ratios.

Table 5 also shows that the sensitivity of the leverage ratio to the primitive parameters varies across the models. At least for the constellation of parameter values we consider, the dynamic leverage ratio is more sensitive to asset volatility and less sensitive to changes in the tax rate and default costs. In all cases, the flexibility to adjust in the future makes leverage more valuable. For example, higher asset volatility introduces substantial kurtosis in the distribution of the future cash flows and therefore makes dynamic adjustment more valuable. The comparison between dynamic and static cases is most vivid in the comparison of their value differentials (relative to the unlevered firm value). In the benchmark endogenous case, adjusting leverage in the future increases the value differential by 80%, and the magnitude varies between an increase by 50% and a triple rise. For

[14]This change explains some minor differences between the otherwise similar results in Tables 4 and 5 for the static model.

example, if the tax benefits are high at the current statutory corporate tax rate of 0.40 (vs. 0.2 assumed in the benchmark case), dynamic flexibility increases the firm value by half, whereas it increases by one-quarter in the static capital structure case.

If the firm defaults the first time its earnings are less than or equal to coupon, the comparison between the dynamic and static models remains similar, but the leverage ratios are substantially smaller. In the benchmark case, the leverage ratio changes from 52% to 17% with a change in the default policy, and the value differential is substantially smaller.

2.4.4 Discussion

The basic and first-order quantitative implication of allowing firms to refinance in response to an increase in future profits is a significant decrease in model-implied leverage ratios at the time of debt issuance, bringing them close to the leverage ratios observed in reality. Note that the importance of future profitability, an intuitive and economically plausible explanation, is difficult to study even qualitatively (not to mention quantitatively) using conventional static methods. The extension of the contingent claims paradigm to dynamic leverage policies allows us to model the impact of future changes in leverage policy in response to profitability shocks on current financial decisions. It also enables us to measure the extent to which future interactions explain empirically observed leverage today. This common theme persists across all the applications of this paradigm.

The impact of introducing dynamic refinancing on leverage ratios has been discussed by Fischer, Heinkel, and Zechner (1989) and Goldstein, Ju, and Leland (2001). The basic finding of Goldstein, Ju, and Leland (2001) is that when firms are allowed to adjust their leverage dynamically, the optimal leverage at the initial date changes (for their benchmark set of parameter values) from 50% to 37%.[15]

Quantitative results also confirm the extent of the adjustment cost effect, which is very similar

[15] The leverage ratio of 50% in their base case (i.e., when firms cannot adjust dynamically) is lower than the leverage of approximately 70% in our static endogenous default model — mostly because of different tax assumptions. That is, Goldstein, Ju, and Leland (2001) assume that firms partially lose their tax benefits when their earnings are sufficiently low. Although we do not detail the economic reasoning for these assumptions here, we note that their usage illustrates once again that the introduction of new variables or economic mechanisms can be studied within the same framework, and their importance can be assessed quantitatively.

in nature to that in menu cost models. Even though the transaction costs of debt issuance are relatively small (in our benchmark case, only 1% of the newly issued debt), their impact on both optimal leverage and, as we shall see, the expected time between refinancing dates is substantial. In particular, small adjustment costs produce large inaction regions where firms "passively" observe changes in their fortunes and leverage without actively adjusting their external financing. Indeed, in the model we have developed, firms are frequently inactive. Yet, when firms finally refinance, they do so in a lumpy fashion, issuing a lot of debt (in addition to the debt they have recalled under the model assumptions). At the heart of the matter is the realization that this "passive" behavior is entirely optimal and rational.

The observation that adjustment costs lead to lumpy behavior is not new. For example, research has extensively investigated the relationship between adjustment costs and lumpiness in the context of investment, hiring, and wage setting (Stokey (2009) discusses these and many other examples). Studies in various asset pricing contexts has been exploring the impact of transaction costs for a long time as well. However, inaction and lumpiness have entered corporate finance research relatively lately, perhaps because any study of such issues must involve a fully fledged dynamic model. Thus, an example of capital structure serves not only as a means to address an interesting research question in its own right, but also as an illustration of a wider agenda and methods that can be applied to many other topics studied under the corporate finance umbrella.

2.5 Cross-Sectional Implications of Dynamic Capital Structure

How well do these models account for the empirically observed behavior of firms? A model with static capital structure decisions produces leverage that is much too high to be consistent with what is observed in reality. Allowing firms to adjust debt levels dynamically results in much lower leverage. Dynamic models thus give rise to a hope that dynamic adjustment can facilitate a reconciliation of the model with the data.

There is, however, a substantial glitch in this result, as identified first in a corporate finance context by Strebulaev (2007). He emphasizes that findings about "optimal" capital structure relate only to those moments in time when "active" financial decisions are made and leverage is actively adjusted. They thus cover only one set of observations. Indeed, as these models predict, and as empirical evidence suggests, firms do not adjust continuously. Instead, empirical data on leverage come from the cross-section of firms that spend substantial time in the inaction region between consecutive refinancings. Therefore, it is inappropriate to compare empirical cross-sectional leverage ratios with the model-implied leverage ratios *at refinancing dates* because they typically refer to different points in time. The extent of the bias depends on the properties of the cross-sectional distribution of leverage. To draw comparisons between dynamic models and empirical data, we need to make sure that the data implied by the model and the data taken from real firms have mutually consistent structures.

Such an idea is obviously inspired by dynamic considerations, and simply could not have been entertained within the static paradigm. As a practical implementation of this idea, it is important to determine the extent to which cross-sectional leverage differs from leverage at times of adjustment. Such a question can be answered only by quantifying the dynamic intuition using a formal model. To this end, Strebulaev (2007) simulates leverage dynamics for the cross-section of firms implied by a dynamic capital structure model. If the model represents the true financial policies of firms in the real world, then the cross-sectional data it produces should be similar to the cross-sectional data available to empiricists. as in Bhamra, Kuehn, and Strebulaev (2010a), we refer to the dynamics of the entire cross-section as "true dynamics," to distinguish these dynamics from the "dynamics" pertaining only to times of refinancing. Although the model describes the capital structure decisions of an individual firm, nothing prevents us from using it for a cross-section of firms, assuming that all firms make optimal decisions independently of others.

2.5.1 Leverage in True Dynamics vs. Leverage at Refinancing Points

Consider an economy consisting of N firms, each of which is driven by the dynamic capital structure model in Section 2.4. Assume that all firms are *identical* replicas of one another at the time of refinancing.[16] Although their optimal leverage at refinancing is thus the same, their leverage at any other time could be quite different if idiosyncratic shocks hitting firms' fortunes are not perfectly correlated. For illustration, let's assume that the Brownian components of firms' cash flow stochastic processes are independent, such that all random shocks are idiosyncratic.

In this case, we are interested in the dynamic behavior of only one variable of interest, the leverage ratio L, which for the dynamic model at any point in time between refinancing waves is expressed as $L_t = \frac{D(X_t)}{D(X_t)+s(X_t)}$, where $D(X_t)$ is given by Equation (2.26) and $s(X_t)$ by Equation (2.28). We resort to a simulation of this model, though recently Morellec, Nikolov, and Schürhoff (2012) and Korteweg and Strebulaev (2012) investigated similar models' performance by deriving the cross-sectional stationary distributions of leverage. Simulation has the advantages of generality, because it can be applied equally well to more complicated models and, in our case, simplicity.

Here is how the simulation works: At date 0, N all-equity firms in the economy are "born" and raise debt for the first time. All firms at that point are identical. Every period after that, shocks to their cash flows are realized (in our case, only idiosyncratic shocks), and if neither the default nor the refinancing thresholds are reached, firms optimally take no actions. It is important to note that though the valuation takes place under the risk-neutral distribution, actual shock realization falls under the actual distribution, so the risk premium needs to be taken into account.[17] If the refinancing trigger is activated, firms recall existing debt and issue new debt. Because we care only about the leverage *ratio*, we assume without loss of generality that at that moment the firm is "re-born" anew, such that all its cash flows are normalized to X_0. If the firm defaults, a new firm replaces the defaulted one and raises debt for the first time (this simplifying assumption keeps the number of firms in the economy constant). We continue the process for T_0 years, where T_0 is large enough to ensure the economy converges to its stationary cross-sectional distribution. Then we simulate the economy for T further years and study the resulting $T \times N$ panel data set. For

[16] Note that in the model we have developed, size is irrelevant because of the first-order homogeneity property. Kurshev and Strebulaev (2006) introduce truly fixed costs of debt issuance that make size an important factor in decision making.

[17] The risk premium accounts for the difference between risk-neutral and actual processes. In the case of Brownian shocks, the difference is in the instantaneous growth rate.

this example, we choose T_0 of 150 years, T of 30 years, N equal to 5,000, and ΔT, which is the length of one period in a simulation, of one quarter.

Table 6 reports the results of this exercise for both endogenous and exogenous default cases, where the benchmark set of parameters is the same as in Section 2.4. In addition, the risk premium is assumed to be 5%. The first two columns present the optimal date-0 leverage ratios in the static and dynamic models for comparison. The next three columns summarize the cross-sectional behavior of leverage in true dynamics. Consider the endogenous default case. For the benchmark dynamic capital structure, the date-0 leverage is 0.52. The firm refinances when its leverage decreases to 0.35 and defaults when its leverage reaches 1. The distribution of leverage in true dynamics thus falls between 0.35 and 1. As expected, there is substantial variation in the cross-sectional distribution of leverage, but the average leverage in true dynamics is higher than that at date 0.

What can explain the result in which the mean leverage in true dynamics is not equal to the optimal leverage ratio at refinancing? Figure 9 shows the stationary cross-sectional distribution of leverage and clearly reveals that the distribution is very asymmetric, with the optimal leverage ratio as the mode rather than the mean of the distribution. As Korteweg and Strebulaev (2012) show for a more general case, this result is generic. Other than in special circumstances, the mean of such a distribution does not equal its mode. Correspondingly, Table 6 shows that the average leverage in true dynamics never equals the date-0 leverage ratio for all the parameter sets reported; sometimes, the difference is quite large.

The mean is more likely to be larger than the mode if the date-0 leverage ratio is lower or if the distribution has fatter tails. Thus, when asset volatility is low or the issuance cost is large, the mean is more likely to be lower than the mode. If the date-0 optimal leverage ratio is lower than this model suggests (e.g., if we introduce asymmetric taxation of corporate profits and losses), the difference between the mean and the mode increases. As soon as we realize that the average leverage in true dynamics is what we need to compare against the empirically observed average leverage if we want to be consistent, we also recognize that the model can actually be further away from accounting for stylized leverage facts.

The intuition behind these results is, of course, very general and not confined to a specific dynamic capital structure model. For example, firms are not identical even at refinancing, because they have different characteristics. In addition, their shocks are partially correlated as a result of the systematic macro- and industry-wide shocks that affect the entire economy. Strebulaev (2007) takes these factors into consideration and finds that the average leverage ratio in the dynamic cross-section is typically substantially *higher* than that at refinancing for many sets of parameters. In other words, "true dynamics" largely cancel out the decrease in optimal leverage due to a dynamic adjustment.

2.5.2 Empirical Capital Structure Studies

To explore the power of true dynamics further, we apply a cross-sectional approach, which we have used thus far to understand average leverage ratios, to bridge the gap between theory and traditional empirical studies of capital structure. We start by painting a broad picture of how conventional empirical studies of capital structure work. Early empirical studies of corporate capital structure, such as the well-known and careful study by Titman and Wessels (1988), tried to identify factors that, according to various capital structure theories, should affect corporate financial policies, and then attempted to pinpoint their explanatory power and sign. What Titman and Wessels (1988) and the many empirical studies they inspired all found, however, is that leverage is robustly correlated only with a small set of explanatory variables, such as size, profitability, and the market-to-book ratio. Thus, an important problem is that this approach typically lacks sufficient power to distinguish between competing theories of capital structure, because these hypotheses actually provide the same qualitative conjectures with respect to most of the variables of interest.

A prominent example is a long-standing effort to distinguish between the trade-off theory of capital structure (the dynamic capital structure model that we have developed here is an example of such a theory) and the so-called pecking order model (Myers and Majluf 1984), which is rooted in the idea that there is an inherent information asymmetry between corporate insiders and outsiders. Although the dynamic version of the pecking order theory has not yet been formalized, many of its implications might be derived intuitively. As a good illustration of this pure intuitive approach, Fama and French (2002) collect most of the knowledge and methods accumulated by traditional empirical approaches up to that point. Fama and French (2002) offer an an entirely intuitive, qualitative discussion of the comparative statics implications of static and dynamic versions of the pecking order and trade-off models. In particular, they assert that the effect of profitability is perhaps the only really outstanding difference between the two. In the pecking order, higher profitability leads to less reliance on debt, because internal funds crowd out external debt financing, and thus lower leverage. The trade-off argument instead intuitively calls for higher leverage, because higher profitability increases the potential tax benefits of debt and reduces expected bankruptcy costs. Using a regression analysis, Fama and French (2002) find, consistent with most prior research, that leverage is negatively related to profitability and thus conclude that this is "the important failure of the trade-off model" (p. 29). Indeed, this finding is identified as *only* "*one* scar on the trade-off model"(p. 30, added emphasis).

The shortcoming of this intuitive approach should by now be clear. Most authors simply confuse the predictions of the dynamic trade-off model at the point of refinancing and "true dynamics." If our goal is to compare model predictions with the results of a conventional regression analysis, we should consider only the latter. Hence, these studies cannot hope to measure the extent to which the trade-off model actually explains the real behavior of firms in an economy. The dynamic relation between leverage and profitability is a particularly striking example of a compelling test of the credibility of empirical cross-sectional research.

To do so, we use the same procedure that we adopted to analyze average leverage in the dynamic cross-section. Profitability is the ratio of earnings before interest and taxes (EBIT, or the sum of the net payout and retained earnings) to the book value of assets in place. In this model context, X_t is the net payout to claimholders, and retained earnings refer to the change in the value of book assets. Because the investment process is fixed, we assume that the book assets grow at a rate equal to the growth rate of the net payout. This growth rate is the only one, for which the market-to-book ratio for assets has a finite nonzero expected value at the infinite time horizon.

Table 7 reports the results of a similar experiment from Strebulaev (2007) (see Table IV, p.

1771). The regressions, consistent with empirical studies, focus on quasi-market leverage (see Equation (2.22)), and the experiment introduces heterogeneity across firms at refinancing points by varying primitive parameters and thus including several controls (e.g., asset volatility, default costs). It also introduces systematic risk and therefore runs simulations for many economies (for details see Strebulaev (2007)). The last two columns show the 10th and 90th percentile values of regression coefficients across economies. The empirical methodology matches several prominent papers, though we report only the results from running the Fama and French (2002) regressions on the model-simulated data.

It is important to understand the intuition behind these results: When all firms reach the refinancing date, the regression of leverage on profitability produces a result that, not surprisingly, matches the intuition proposed by Fama and French and others, namely, that leverage and profitability are positively related in the world described by the trade-off model. This result is shown in the first column of Table 7. However, the same regression performed on both actively rebalancing and passive firms produces a *negative* relation between leverage and profitability, in line with previous empirical findings. (A recent empirical work by Danis, Rettl, and Whited (2012) comes to similar conclusions.) An important implication is that empirical researchers, such as Fama and French, review the data simulated by this model, they would likely interpret the findings as evidence in favor of the pecking order argument, contrary to the predictions of the dynamic trade-off model, even though of course the data set was produced by a trade-off model in the first place! In short, as Strebulaev (2007) shows, this, and likely many other quantitative, as wells some qualitative, conclusions of extant literature are unwarranted, mostly because the studies that achieve these conclusions ignored true dynamics.

The negative relationship between leverage and profitability might instead be understood in the modeling context by taking dynamic considerations into account. Profitability is persistent, so an increase in profitability affects future profitability and thus the value of the firm. Both the market value, through embedded expectations, and the book value, through retained earnings, increase. But whereas an increase in the value of the firm always lowers leverage, it does not necessarily lead to refinancing in a world with infrequent adjustment. Debt levels, therefore, are not affected unless a sufficiently high profitability level is reached, triggering refinancing.[18] Similar considerations hold as profitability decreases. As a result, for any individual firm between two refinancing dates the relationship between leverage and profitability is strictly negative. The positive relationship, on the other hand, is an outcome of cross-sectional forces at the refinancing date. Although these arguments seem intuitive, assessing their quantitative relevance requires a dynamic model. The simulations based on such a model clarify the extent to which the negative relation between leverage and profitability dominates the positive relation we can observe at refinancing. The result on the prevalence of the negative effect for a wide range of parameters is implied by a quantitative assessment of the dynamic behavior of the model, not on the qualitative comparative statics. As Strebulaev (2007) shows, the quantitative values of the profitability slope coefficient as reported by Fama and French (2002) are consistent with the coefficient values produced in his model-implied simulations. In other words, empirically observed profitability coefficients are generated within a reasonable range of simulated economies.

Overall then, dynamic cross-sectional methods applied to a dynamic trade-off contingent claims model indicate that the most important and seemingly non-controversial result in empirical capital structure work in fact offers little if any value.

2.6 Macroeconomic Fundamentals and Capital Structure

How do macroeconomic factors affect corporate financial decisions? How do corporate financial decisions by individual firms affect the aggregate properties of the economy? Because dynamic contingent claims models are closely related to asset pricing, they offer us an opportunity to consider issues that traditionally have been the preserve of asset pricing, such as the equity premium puzzle. These models can address asset pricing issues at the same time as corporate finance issues by successfully blending contingent claims models and traditional asset pricing models.

The rationale for pursuing this research is straightforward. Consider traditional consumption-

[18] To be precise, the book debt values are not affected. The sensitivity of the market value of debt to profitability is much smaller than that of equity value, and if we replace quasi-market leverage with market leverage, the results are barely affected.

based asset pricing models, starting with Lucas (1978), whose goal is to understand various quantitative features of the macroeconomic data, such as the equity premium puzzle of Mehra and Prescott (1985). These models are typically built around a representative agent, who acts as a consumer as well as the owner of financial assets in the economy. However, corporations live in the same world, and firm owners and managers are also consumers. A growing body of empirical work indicates that common factors may affect the equity risk premium and credit spreads on corporate bonds. The economic mechanisms behind these "puzzles" (e.g., equity premium, credit risk, risk-free rate puzzles) are likely linked to the economic mechanisms underlying optimal financial decisions.

This intuition forms the premise for work by Bhamra, Kuehn, and Strebulaev (2010b) that builds on the framework developed by Bhamra, Kuehn, and Strebulaev (2010a) and earlier contributions by Hackbarth, Miao, and Morellec (2006) and Guo, Miao, and Morellec (2005). A similar set of issues is independently considered by Chen (2010). This stream of research develops a new unified framework that explores asset pricing and corporate finance issues *simultaneously* and thereby enables us to study how the optimal financial structure of firms, at both individual and aggregate levels, is affected by time-varying macroeconomic conditions. Therefore, in this section, we briefly introduce and discuss this theoretical framework at an intuitive level (a more extensive formal development requires additional technical apparatus, which is beyond the scope of this review). We also discuss how this framework might reconcile existing empirical evidence about the relationship between capital structure and macroeconomic factors. Empirically, aggregate leverage in the economy appears counter-cyclical (e.g., Korajczyk and Levy (2003)), yet parallel evidence suggests that firm-level leverage is pro-cyclical (e.g., Korteweg (2010)).

In a nutshell, Bhamra, Kuehn, and Strebulaev (2010b) embed a dynamic capital structure model of the type described in Section 2.4 within a dynamic consumption-based asset pricing model with a representative agent. From an economic perspective, this approach enables us to study the impact of macroeconomic factors on the behavior of individual and aggregate corporate capital structure. The advantage of this type of embedded model over a pure contingent claims model is that it endogenously links unobservable risk-neutral probabilities with observable actual probabilities through the market price of consumption risk and the agent's preferences. In turn, we can investigate the effect of macroeconomic risk on optimal financing decisions endogenously, because the agent prices the claims on the basis of his or her attitude toward the macroeconomic risks present in consumption and cash flows. With this framework, we also explain intuitively how preferences (e.g., risk aversion, the elasticity of intertemporal substitution) affect financing decisions. Finally, the framework enables us to study actual default probabilities and, in particular, explore the relation between leverage and default likelihood. All these tasks would have been impossible with the previous generation of contingent claims models, because the pricing kernel is given exogenously, and so there are no macro factors to speak of.

The economy consists of consecutive periods of stochastic booms and contractions that affect, through consumer preferences, asset prices and optimal corporate decisions. Business cycles are modeled as a two-state Markov chain process, where the probability of switching from one state into another is asymmetric in such a way that agents expect the economy to spend more time in booms than in contractions. Booms differ from contractions along several dimensions: The growth rate (volatility) of the cash flow and consumption processes is higher (lower) in booms, whereas the correlation between the firm's cash flows and consumption is higher in contractions.

One significant consequence of financial decisions in this model is that optimal leverage at refinancing is now state-dependent: If firms refinance in bad times, they optimally choose lower, leverage because cash flows are expected to be lower in the future, and the distance to default is expected to be shorter. For a similar reason, firms are more likely to default in contractions, all else being equal. This path dependence, however, leads to a more intricate relation between macroeconomics and leverage. Firms that lever up in expansions have higher leverage and are more vulnerable in recessions. As an interesting empirical prediction, we note that firms that choose to raise debt in booms are more prone to default when macroeconomic conditions worsen. From an aggregate perspective, optimal levering up in booms leads to default clustering in contractions. Conversely, firms that refinance in recessions are more likely to be relatively underlevered in booms and thus enjoy lower tax benefits from issuing debt. They also are more likely to refinance in an expansion. These and many other predictions of the framework have not yet been tested empirically.

As the trade-off intuition implies, optimal leverage is pro-cyclical at the refinancing point. However, aggregate leverage is actually counter-cyclical in true dynamics. The direction of cyclicality differs between true dynamics and refinancing points, because when the state of the economy worsens, the market value of equity falls more than the market value of debt, for equity, as a residual claim, is more sensitive to macroeconomic risk. This mechanism induces counter-cyclicality in leverage. In true dynamics, the latter effect dominates the optimal choice of debt at refinancing. The difference between true dynamics and refinancing points explains why it is not an easy task to study the relation between macroeconomic conditions and financial structure in the data and why various empirical studies achieve seemingly inconsistent results. The result related to the counter-cyclicality of the whole cross-section is consistent, for example, with Korajczyk and Levy (2003), who find leverage to be counter-cyclical for relatively unconstrained firms. Yet, Covas and Den Haan (2007) and Korteweg (2010) empirically find that when firms choose optimal leverage, leverage decisions are pro-cyclical, consistent with the refinancing date result. The intuition behind this result is similar to our approach in Section 2.5, where we contrast the refinancing point (or date-0) leverage ratio and the cross-sectional leverage ratio in true dynamics.

The framework obviously delivers multiple interesting quantitative results. Macroeconomic risk leads to substantially lower leverage at refinancing. Under the benchmark set of parameters considered by Bhamra, Kuehn, and Strebulaev (2010b), the date-0 leverage ratio is 32% in a boom and 22% in a contraction. The unconditional (i.e., independent of the state of the economy) leverage ratio is 28%, consistent with stylized facts pertaining to the extent of debt usage. In true dynamics however, consistent with the results of Section 2.5, leverage is substantially higher in both bad and good times, rising from 22% to 44% in contractions and from 32% to 38% in booms.

Of course, Bhamra, Kuehn, and Strebulaev (2010b) and others also emphasize how much we do not yet know about the relation between macroeconomics and corporate financial and investment decisions. For example, the model assumes that consumption is unaffected by corporate defaults and that, as in traditional consumption-based models, all decisions are made by a representative agent. Intuitively, both investment and consumption may interact with corporate financial policies. Also, financial claims, such as corporate debt and equity, are held by different agents, which may prompt conflicts of interest and important implications for default and leverage. Above all, we do not know how these factors affect leverage dynamics.

2.7 Debt Structure and Strategic Renegotiations

Our treatment of debt has so far assumed that if the firm is unable to cover its debt obligations, the ensuing default leads either to a reorganization or a bankruptcy. In other words, we have assumed away the possibility of renegotiation. This scenario is realistic in the case of a firm issuing public bonds under the U.S. jurisdiction, because, according to the 1939 Trust Indenture act, any change in the major features of a debt contract, such as principal, interest, or maturity, require a unanimous consent of all bondholders, a practically infeasible task. However, U.S. public bond markets represent only one side of the debt market. Another is represented by private debt contracts, such as bank debt. A critical difference between public and private debt is that firms can attempt to renegotiate the latter and, if the renegotiation is successful, the firms change the structure of payments to debtholders to avoid default. This gives rise to shareholders behavior known as strategic renegotiation: shareholders refuse to honor debt obligations in full and threaten default, even though they have sufficient resources to fulfill their promises. Such a threat can be effective only if debtholders face default costs they prefer to avoid.

The importance of strategic renegotiation was explored originally in the credit risk literature, pioneered by a classical paper of Anderson and Sundaresan (1996) and followed by Mella-Barral and Perraudin (1997) and Fan and Sundaresan (2000). We consider here a simple version of a static capital structure model with strategic renegotiation, closely following Hackbarth, Hennessy, and Leland (2007). The model is similar to the model of Section 2.3 (the case of endogenous default with liquidation), but in addition to public debt, private (bank) debt can be issued by the firm. The only difference between public and bank debt is in the possibility that the firm can renegotiate bank debt at any time

First consider the scenario when only bank debt with principal B is outstanding and the firm has a strong bargaining position, i.e., it has the ability to make take-it-or-leave-it offers to the bank. If the bank rejects the offer, the firm defaults and the firm is liquidated. In this case, the bank's promised payoff is given by:

$$R_B(X) = \min\left[\frac{c_B}{r}, (1-\alpha)(1-\tau)\frac{X}{r-\mu}\right], \tag{2.32}$$

where c_B is the initially promised coupon, and $(1-\alpha)(1-\tau)\frac{X}{r-\mu}$ is the post-bankruptcy value, which is similar to that given by Equation (2.9). R_B thus represents the bank's reservation value, because the bank would reject any offer that yields a lower payoff than R_B. If the shareholders have all the bargaining power, the bank's promised claim always equals R_B. In the absence of public non-renegotiable debt, the firm never defaults but rather negotiates its debt payments in bad states of the world. In this model, the firm renegotiates its debt if the value of cash flow reaches some value X_S for the first time. Because the firm has full bargaining power, the firm renegotiates to keep the value of bank debt at its reservation value $(1-\alpha)(1-\tau)\frac{X}{r-\mu}$. As a result, the firm pays a lower coupon rate, given by:

$$s(X) = (r-\mu)(1-\alpha)(1-\tau)\frac{X}{r-\mu}. \tag{2.33}$$

Note that the firm never defaults and thus default costs do not occur on equilibrium path. The total firm value can be written as the sum of the after-tax unlevered firm value and the present value of bank debt tax benefits:

$$V(X) = (1-\tau)\frac{X}{r-\mu} + \tau B(X), \tag{2.34}$$

where $B(X)$ is the value of bank debt. Outside of the renegotiation region the value of the bank debt is determined using the same logic we used in Section 2.3:

$$B(X) = \frac{c_B}{r} - \left(\frac{c_B}{r} - (1-\alpha)(1-\tau)\frac{X_S}{r-\mu}\right)\left(\frac{X}{X_S}\right)^{\xi_1}, \quad X \geq X_S. \tag{2.35}$$

The first term is the value of riskless bank debt and the second term takes into account the impact of renegotiation. Equityholders choose X_S by maximizing equity value:

$$E(X) = V(X) - B(X) = (1-\tau)\frac{X}{r-\mu} - (1-\tau)B(X). \tag{2.36}$$

The first order condition with respect to the renegotiation threshold X_S yields:

$$X_S = \frac{c_B}{r}\frac{\xi_1}{\xi_1 - 1}\frac{r-\mu}{(1-\alpha)(1-\tau)}. \tag{2.37}$$

At date 0, maximizing equity value with respect to c_B and taking the threshold value into account, gives the optimal coupon rate:

$$c_B^* = \frac{\xi_1 - 1}{\xi_1} r(1-\tau)(1-\alpha)\frac{X_0}{r-\mu}. \tag{2.38}$$

The optimal levered firm value is then:

$$V(X_0) = (1-\tau)(1+\tau(1-\alpha))\frac{X_0}{r-\mu}. \tag{2.39}$$

Is the optimal leverage or firm value higher in public debt or bank debt cases? The comparison of c_B and c (Equations (2.38) and (2.20), respectively) shows that there is no unambiguous answer to this important question. In the bank debt case, the leverage ratio is constrained by the extent of default costs. For example, if the asset value is fully destroyed in default ($\alpha = 1$), the firm can always threaten the bank to default and thus no debt is feasible initially ($c_B = 0$). On the other hand, if default costs are very low, the firm does not have any credible threat to renegotiate the terms of the loan and the resulting leverage is very high.

Table 2.7 compares the two cases corresponding to the parameter sets of Section 2.3.5. For the benchmark set, the bank debt coupon is 2.29 as opposed to 1.36 for public debt, and leverage is 0.76 vs. 0.70. For all the parameter variations in the table, the bank debt coupon is higher than the coupon on public debt. The leverage ratio, however, can be lower in the bank debt case, because bank debt leads to lower ex-ante inefficiency and thus higher total value, fully appropriated by equityholders. The extent of an increase in value is gauged by the ratio of levered to unlevered firm values, which is substantially large for the bank debt case.

So far we have assumed that firms can use only one source of financing, although empirically many firms issue both bank and public debt at the same time. To incorporate two types of debt in the same model we need to decide on the seniority of debt claims. Hackbarth, Hennessy, and Leland (2007) show that in this model it is optimal to give a strict seniority to bank debt, consistent with empirical evidence. Bank debt seniority protects the firm from ex-ante inefficient renegotiation if the cash flow level is relatively high and facilitates ex-post efficient renegotiation in the bad states of the world.

The renegotiation region is modified to take into account the endogenous default on public debt, which occurs when the cash flow level reaches X_D. The renegotiation region is then $[X_D, X_S)$ and the non-distress region is $[X_S, \infty)$. For simplicity, assume that the recovery of public debtholders in default is zero. Because the presence of public debt does not change the renegotiation trigger, the solution for X_S is the same. The value of public debt can then be written as:

$$D(X) = \frac{c}{r}\left(1 - \left(\frac{X}{X_D}\right)^{\xi_1}\right). \tag{2.40}$$

The smooth-pasting condition gives the following default threshold:

$$X_D = \frac{\xi_1}{\xi_1 - 1}\frac{c}{r-\mu}r(1-(1-\tau)(1-\alpha)). \tag{2.41}$$

Firm's objective function at date 0 includes all the benefits and costs of both types of debt:

$$V(X_0) = (1-\tau)\frac{X_0}{r-\mu} + \tau B(X_0) + \tau D(X_0) - (1-\tau)\frac{X_D}{r-\mu}(1-(1-\tau)(1-\alpha))\left(\frac{X}{X_D}\right)^{\xi_1}. \tag{2.42}$$

As before, the value function is increasing in the value of bank debt and thus the optimal bank coupon rate is still defined by (2.38). Optimization with respect to public debt coupon yields:

$$c^* = \frac{\xi_1 - 1}{\xi_1}\frac{r}{r-\mu}\left(\frac{\tau}{\tau-\xi_1}\right)^{-\frac{1}{\xi_1}}(1-(1-\alpha)(1-\tau))X_0. \tag{2.43}$$

Note that public debt is a complement rather than a substitute for bank debt, because the firm still wants to issue as much bank debt as possible. The last set of columns in Table 2.7 shows that the additional public coupon is 0.16 in the benchmark case, with the leverage ratio increasing to 0.79. In this version of the model, however, adding public debt makes little quantitative impact on the total extent of debt.

2.8 Capital Structure and Temporary vs. Permanent Shocks

Empirical evidence strongly suggests that financial managers consider financial flexibility, earnings and cash flow volatility, and insufficient internal funds the most important determinants of financing decisions by their firms (e.g., Graham and Harvey (2001)). Although intuitively appealing, these factors continue to puzzle corporate finance theorists, who find it hard to pin down the exact mechanism of their influence. We may know the nature of economic frictions that push firms to adopt more cautious financing behavior, but existing attempts to quantify their effects leave a great deal to be desired. In Section 3, we introduce endogenous investment models that analyze these issues directly. Here, we consider a paradigm change within the dynamic contingent claims models that takes these issues into account.

2.8.1 Impact of Temporary Shocks

If we reconsider the dynamic capital structure model of Section 2.4, we confront some difficulty explaining these stylized facts. In fact, the model induces several undesirable implications. First, the non-negativity of market values implies that cash flows in the model can never be negative, which is, of course, highly unrealistic. For example, between 1987 and 2005, approximately 17% to 25% of all quarterly cash flows for the full Compustat sample were negative. On a broader level, the model ignores the strong likelihood that firms that are highly profitable today may experience negative cash flows in the near future. In other words, firms value financial flexibility as a means to sustain them when cash flows are very low, even though their future is bright. For example, consider the economic marginal tax rate that Scholes and Wolfson (1992) define to account for the present value of current and future taxes on an extra dollar of income earned. Existing models would predict that profitable firms sustain their profitability in the future. They are therefore assigned high tax rates even though in reality their profitability is less persistent and their true marginal tax rate should be lower. Their realistic expectations should lower their inclination to issue debt.

Second, the volatilities of cash flow and asset value growth rates are equal in this model, which makes it difficult to reconcile with the reality in which managers care more about earnings volatility than about asset volatility. Because the asset volatilities of most firms are relatively low, the distinction is an important one.[19] Third, both changes in and levels of current cash flow and asset value are perfectly correlated, which is inconsistent with both intuition and evidence.

As it turns out, these problematic features all result from the permanent nature of cash flow shocks in the model, and this permanence itself is caused in part by reliance on geometric Brownian motion as an underlying cash flow process. Although this process is typically assumed for tractability, its historical significance should not be understated: The direct predecessor of dynamic contingent claims models is the Black-Scholes model, in which the behavior of stock prices (unlike corporate cash flows) is much more similar to the behavior exhibited by geometric Brownian motion.

Investigating the importance of the temporary nature of cash flow shocks is the main focus of Gorbenko and Strebulaev (2010), and we follow their approach closely in this section. [20] Consider the effect of introducing a temporary component to an otherwise standard cash flow process. The exact nature or sources of these shocks is irrelevant, because this intuition is generic. For example, shocks could affect demand for a firm's product or the production cost structure. Because their impact is of limited duration, these transitory shocks affect future cash flows disproportionately over the time horizon. The consequences are many: The total realized current cash flow can turn negative (even if asset values are large), earnings and asset volatilities differ, and levels and innovations in current cash flows are imperfectly (though positively) correlated with those in asset values. Overall then, introducing temporary shocks makes the model's behavior patterns consistent with stylized facts about corporate earnings, along multiple dimensions.

More importantly, we need to determine the contribution of this mechanism to explanations of corporate behavior. Intuitively, for any given level of leverage, the expected advantage to debt is likely of lower value when cash flows are more volatile thus prompt predictions that they will be

[19]For example, Schaefer and Strebulaev (2008) estimate annualized asset volatility for an average firm with an existing investment-grade rating to be approximately 23%.

[20]Grenadier and Malenko (2010) explore the impact of similarly modeled temporary shocks in a real options investment problem.

low more frequently. For example, because we expect cash flow often to be lower than the level of contractual financial obligations, the expected tax benefits of debt should diminish in the absence of full tax loss offset provisions. In other words, incentives to use debt decrease with the probability that a firm will experience nontaxable states of the world (Kim 1989; Graham 2003).

To quantify this intuition, we build a fully fledged dynamic contingent claims model with both permanent and temporary shocks. The main finding is that the value of financial flexibility may indeed increase dramatically when firms face prospects of immediate adverse temporary shocks of large magnitude. For the benchmark set of parameters, the optimal static leverage ratio decreases from 58% to 37% with the introduction of temporary shocks, conditional on retaining the *same* total asset volatility. As expected, the model predicts that the effect of temporary shocks is disproportionately larger for small firms (an almost 30% decrease in leverage, compared with the permanent-only model), consistent with empirical evidence about the relation between firm size and leverage. Temporary shocks appear particularly important when asset volatility is low. For example, when asset volatility is 0.25 per annum (an empirically realistic value), leverage is lower by 27%, but when asset volatility is 0.5 (considerably higher than asset volatility for the average Compustat firm), leverage decreases only by approximately 3%. Financially constrained firms are also more affected by temporary shocks. For example, a firm with a 20% proportional cost of raising money in financial distress chooses an optimal leverage of 31%, compared with a financially unconstrained firm, whose optimal leverage is 47%. The model further predicts that if managers (but not investors) are myopic and take into account only a limited number of shocks, their remaining desire to issue debt rapidly decreases. For the same benchmark set of parameters, if the manager expects only one large shock in the future, optimal leverage falls to 20%, similar to the quasi-market leverage ratio of a typical public U.S. firm.

2.8.2 Modeling Temporary Shocks

Building a framework that fully incorporates both types of shocks entails formidable technical challenges. The lack of perfect correlation between permanent and temporary shocks gives rise to two sources of dynamic uncertainty that make all financing decisions intrinsically complicated.

Consider a firm whose cash flows are driven by two stochastic processes: one responsible for the long-term prospects (e.g., a geometric Brownian motion) and another that constitutes the short-term deviations from long-term cash flow (e.g., any process that mean-reverts to long-term cash flow levels). In the permanent-only model, the default boundary, as we have seen, is a certain cash flow threshold level. But with transitory shocks, the knowledge of either total asset value or total realized cash flow by itself is insufficient to make a well-informed default decision. The current cash flow could be low because the firm's future is bleak (and it is optimal to default). Alternatively, it might be low because the future prospects are bright, conditional on the firm surviving a temporary period of losses, in which case it is optimal to make debt payments. Only by decomposing the shock into permanent and temporary components can equityholders make the optimal default decision. This problem is very difficult to solve, though, because the default boundary is a curve in a two-dimensional permanent-temporary shock space. Mathematical problems related to the first passage to such a boundary have not been solved satisfactorily even for the simplest cases.

Gorbenko and Strebulaev (2010) instead build a simplified framework that incorporates all these intuitive features. To see how their model works, we start by assuming that at any time t, the realized cash flow, π_t, is the sum of two components: the cash flow generated by the permanent component, X_t, which reflects the long-term prospects of project returns, and the mean-reverting cash flow, ϵ_t, which reflects deviations from the permanent component caused by shocks of shorter duration:

$$\pi_t = X_t + \epsilon_t. \tag{2.44}$$

If $\epsilon_t \equiv 0$, and X follows geometric Brownian motion, we are back to the capital structure model of Section 2.3. To illustrate the modeling of the temporary shock, we shut down the permanent component ($X_t = 1$) and allow for only one transitory shock to arrive in the future. Important features of the shock include the random timing of the shock's arrival, the initial magnitude of the shock, and its persistence. Let the shock arrive at a random moment T_s (s refers to the shock's arrival); its size ϵ_{T_s} is also random and determined at T_s. The transitory nature of the shock is reflected in its degree of persistence; that is, the cash flow mean reverts from $X_t + \epsilon_{T_s}$ to X_t at a random time T_r (r indicates the shock's reversal).

Both the timing of the shock arrival T_s and its persistence $T_r - T_s$ follow a Poisson process with intensity λ that we can write through the following probability density function:

$$\phi(w) = \lambda_w e^{-\lambda_w w}, \quad w = \{s, r\}. \tag{2.45}$$

Two apparently similar parameters, λ_s and λ_r, play distinctly different economic roles, such that λ_s reflects the frequency of "big" transitory shocks hitting the firm, and higher values of λ_s imply the firm is more often "in the news." In contrast, λ_r reflects the speed of the shock's mean reversion. Indeed, the expected value of the shock size attenuates exponentially with the speed λ_r after the occurrence date T_s (if the shock has not yet disappeared). At any date t, the *expected* value of the transitory shock to the cash flow process at any future date $\tau > t$ can be written as:

$$\mathbb{E}_t[\epsilon_\tau] = \begin{cases} 0, & t < T_s \text{ and } t \geq T_r, \\ \epsilon_{T_s} e^{-\lambda_r(\tau - t)}, & T_s \leq t < T_r. \end{cases} \tag{2.46}$$

The *realized* transitory shock to the cash flow can similarly be written for any date t as:

$$\epsilon_t = \begin{cases} 0, & t < T_s \text{ and } t \geq T_r, \\ \epsilon_{T_s}, & T_s \leq t < T_r. \end{cases} \tag{2.47}$$

The initial size of the jump is distributed according to the symmetric double exponential distribution with the following probability density function:

$$f(\epsilon_{T_s}) = \frac{1}{2}\gamma e^{-\gamma|\epsilon_{T_s}|}. \tag{2.48}$$

The parameter γ drives the volatility of the temporary shock, such that smaller γ leads to shocks of larger magnitude.

In this model with temporary shocks, the expected cash flow volatility differs from asset value volatility, and the expected time-series correlation between asset value and cash flows is positive but less than 1. Finally, this setup naturally accommodates the possibility of negative cash flows. Even a naive model with temporary shocks thus satisfies all the empirical criteria for a firm's cash flows, as outlined in the introduction.

In the presence of negative shocks, the owners of the firm may find it optimal to abandon the project. Modeling abandonment in this case is similar to its counterpart in real option models. The threshold level of abandonment is denoted by ϵ_A, at which point firm owners are indifferent between abandonment and continuation, and this value can be found from the indifference condition by equating the residual firm value to zero:

$$\mathbb{E}_{T_s}\left[\int_{T_s}^{T_r}(1+\epsilon_A)\,e^{-r(s-T_s)}ds + e^{-T_r}\int_{T_r}^{\infty}e^{-r(s-T_s)}ds\right] = 0. \tag{2.49}$$

The abandonment level that satisfies Equation (2.49) is:

$$\epsilon_A = -\frac{\lambda_r + r}{r}. \tag{2.50}$$

The transitory nature of the shock dictates that the abandonment level is always lower than the negative value of the permanent cash flow, $\epsilon_A < -1$. Shorter-lived shocks yield a lower abandonment threshold, because they are likely to have a marginal impact on asset value. Lower interest rates also lead to less abandonment, because they render the bright future more valuable after the negative shock is over.

We next define asset value as the present value of all future cash flows streaming from an asset's cash flow rights. It is thus the sum of a perpetuity paying a permanent cash flow and the present value of transitory deviations, which equals:

$$\begin{aligned} V_t &= \frac{X_t}{r} + \mathbb{E}_t\int_t^{\infty}e^{-r(s-t)}\epsilon_s ds \\ &= \frac{1}{r} + \frac{1}{2}\frac{\lambda_s}{r+\lambda_s}\frac{1}{\gamma(r+\lambda_r)}e^{\gamma\epsilon_A}, \quad t < T_s. \end{aligned} \tag{2.51}$$

The second term is always positive and reflects the asymmetric option-like nature of limited liability. As we might expect, asset value increases with the variance of the shock's size, the intensity of the shock's occurrence, and its persistence.

As in previous models, it helps to introduce simple Arrow-Debreu-type securities that pay off upon the shock's arrival. The payoff value depends on the region in which the temporary shock falls. Consider a security $A(g(\epsilon_T), \lambda, b)$ that pays $g(\epsilon_T)$ at time T of the shock's arrival if the size of the shock ϵ_T is at least as great as b. These securities do not assume limited liability – their payoffs, and therefore values, can be negative. The necessary securities have triplets of parameters $(1, \lambda_s, -\infty)$, $(1, \lambda_s, b)$, and $(\epsilon_{T_s}, \lambda_s, b)$. The first one pays \$1 at time T_s for any value of the shock. The second and third securities pay correspondingly \$1 and ϵ_{T_s} if the size of the shock is at least as great as b.

The values of these securities at time $t,\ t < T_s$, are as follows:

$$
\begin{aligned}
A(1, \lambda_s, -\infty) = \mathbb{E}_t\left[e^{-r(T_s-t)}\right] &= \frac{\lambda_s}{r+\lambda_s}, && (2.52)\\
A(1, \lambda_s, b) = A(1, \lambda_s, -\infty)\mathbb{E}_t\left[1|\epsilon_{T_s} > b\right] &= \frac{\lambda_s}{r+\lambda_s}(1 - \frac{1}{2}e^{\gamma b}), && (2.53)\\
A(\epsilon_{T(\lambda_s)}, \lambda_s, b) = A(1, \lambda_s, -\infty)\mathbb{E}_t\left[\epsilon_{T_s}|\epsilon_{T_s} > b\right] &= \frac{\lambda_s}{r+\lambda_s}\frac{1}{2\gamma}e^{\gamma b}(1-\gamma b). && (2.54)
\end{aligned}
$$

The default boundary, to be optimally determined by equityholders is denoted by ϵ_D. The equity value at any time t before the shock occurs can be written as:

$$
E_t = (1-\tau)\left(\frac{1-c}{r}(1 - A(1, \lambda_s, -\infty)) + A(1, \lambda_s, \epsilon_D)\frac{1-c}{r} + A(\epsilon_{T_s}, \lambda_s, \epsilon_D)\frac{1}{r+\lambda_r}\right). \quad (2.55)
$$

The last term in Equation (2.55) is the present value of short-term deviations, which is strictly positive given equity's limited liability. The expected shock thus has an impact on equity value opposite that on debt value: Equity value increases with the shock's frequency, persistence, and magnitude.

In this simple case, the default boundary is determined by the equityholders' indifference condition when the value of equity is zero:

$$
E_{T_s} = \mathbb{E}_{T_s}\left[\int_{T_s}^{\infty}(1 + \epsilon_D \cdot \mathbb{I}_{\{s<T_r\}} - c)e^{-r(s-T_s)}ds\right] = 0, \quad (2.56)
$$

which gives us the optimal default boundary ϵ_D as a function of the coupon rate:

$$
\epsilon_D = -(1-c)\frac{\lambda_r + r}{r}. \quad (2.57)
$$

Similarly, we can write the debt value as:

$$
\begin{aligned}
D_t &= \frac{c}{r}(1 - A(1, \lambda_s, -\infty)) + A(1, \lambda_s, \epsilon_D)\frac{c}{r} && (2.58)\\
&\quad + (1-\alpha)(1-\tau)\Big((A(1, \lambda_s, \epsilon_A) - A(1, \lambda_s, \epsilon_D))\frac{1}{r} + (A(\epsilon_{T_s}, \lambda_s, \epsilon_A) - A(\epsilon_{T_s}, \lambda_s, \epsilon_D))\frac{1}{r+\lambda_r}\Big).
\end{aligned}
$$

The first term reveals the riskless bond value if the shock does not happen. The second term shows that in the case of no default at the time of the shock, the bondholders are entitled to a perpetuity (because it is a single-shock economy). The last term shows that in the case of bankruptcy, the value stems from the perpetuity generated by the permanent cash flow (first component of the third term) and the present value of deviations from the long-term cash flow. To understand the intuition behind this representation, we can think of the complex security $A(\epsilon_{T_s}, \lambda_s, \epsilon_A) - A(\epsilon_{T_s}, \lambda_s, \epsilon_D)$ as the present value of the size of the shock if debtholders incur all of the shock's cost (i.e., the shock is large enough to default but not severe enough to abandon the firm). Debtholders then must pay this present value until the shock mean reverts, as taken into account by the perpetuity formula $\frac{1}{r+\lambda_r}$.

Because default and abandonment can happen only in response to negative shocks (i.e., ϵ_B and ϵ_A are less than zero), the value of $A(\epsilon_{T_s}, \lambda_s, \epsilon_A) - A(\epsilon_{T_s}, \lambda_s, \epsilon_B)$ is negative, and therefore the debt value is *always lower in the presence of temporary shocks* for a given coupon rate and other model parameters. This is a fundamental observation: the total value of the firm is *larger* in the presence of temporary shocks. This apparent inconsistency results from an ex-post conflict of interest between equity and debt. Debtholders incur costs in bad states of the world but do not share in the benefits from good states. It then follows naturally that debt value is a decreasing function of a temporary shock's frequency, its persistence, and its magnitude.

Gorbenko and Strebulaev (2010) extend this simple model in two obvious directions: first, X_t follows the geometric Brownian motion, and, second, many temporary shocks are allowed. What sets the model apart from all previous contingent claims models is that the interest coverage ratio (ratio of earnings to interest payments) and the leverage ratio are two distinct economic measures of a firm's financial health. The current level of cash flow can be low because of a negative temporary shock, which would imply a low interest coverage ratio and potential liquidity problems, whereas the future cash flows can be expected to be high, implying a low leverage ratio. This expectation can explain at least two important stylized facts. First, managers care about cash flow or earnings volatility, instead of asset volatility. Second, firms that default can be broadly allocated to two types, either those with very low net worth (poor future prospects) or those with very low or negative interest coverage ratios (liquidity distress), as described by Davydenko (2010), for example. Overall, accounting for short-term shocks appears crucial for understanding the financial decisions of corporations.

2.9 Further Applications

Recent research has explored a number of further applications of this framework to corporate finance that enrich it along various dimensions. Here, we briefly discuss several of them, which we believe offer substantial promise for the future.

Dynamic Investment and Financing. One drawback of optimal capital structure models that we have considered so far is the absence of a non-trivial investment policy. The separation of investment and financing means that cash flow generating machines are assumed to exist independently of financing considerations. Recent research has started exploring the dynamic interactions of financing and investment policies within the dynamic contingent claims framework. For example, Boyle and Guthrie (2003) enrich the real options model of the type we consider in Section 2.2 by imposing future financing constraints. Intuitively, if in the future the firm is unable to exercise its real option and invest at the optimal time, because funding is unavailable or costly, the real option value is lowered and the firm will tend to exercise earlier. As a result, higher cash flow uncertainty that delays investment in a standard real option setup may have the opposite impact by accelerating investment, because the window of financial opportunities become more valuable.

Campello and Hackbarth (2012) incorporate financial considerations into the standard real options setup by adding costly equity financing and renegotiable bank debt (with a strategic renegotiation structure similar to the one in Section 2.7). Firms that face higher cost of equity issuance invest optimally in projects with lower bankruptcy costs to reduce future renegotiation inefficiencies. Hackbarth, Mathews, and Robinson (2012) analyze the boundaries of the firm by exploring the tension between assets in place and growth options in the presence of the trade-off model of capital structure. In their model, shareholders have access to both assets in place, which produce cash flows as in the capital structure models of Sections 2.3 and 2.4, and a growth option, similar to the model in Section 2.2. However, the exercise of the growth option cannibalizes the cash flow potential of assets in place. The firm can initially choose an organizational form, either by combining the assets in place and the growth option in one firm or by establishing two different firms, with one owning the assets in place and the other having the right to any cash flows from the future exercise of the growth option.

The trade-off faced by the original founders face emphasizes the fundamental tension between substitutes and complementarities. Having one firm internalizes the decision to exercise the real option and thus protects assets in place. In the two-firm setup, exercise of the growth option does not take into account assets in place. Such an exercise strategy maximizes financial flexibility but may be suboptimal for the founding shareholders. This setup can be useful, for example, in studying innovation. Innovative products or business processes negatively affect the profitability of products and processes that they compete with. The framework enables us to explore how the decision-maker's objective affects optimal organizational form (see also Grenadier and Weiss (1997)).

A number of further studies explore such issues, which include the impact of agency conflicts on the interaction of investment and financing (Childs, Mauer, and Ott 2005), growth options and priority structure of debt (Hackbarth and Mauer 2012), undiversified entrepreneurial finance and capital structure in incomplete markets (Chen, Miao, and Wang 2010), and the interaction between time-to-build investment lags and capital structure (Tsyplakov 2008). In a study that merges strategic renegotiation and investment, Pawlina (2010) shows that an ability to renegotiate debt increases the investment trigger and thus exacerbates underinvestment problem.

Asset liquidity. Firms in distress have a number of tools in their arsenal that they can use to avoid default. Models that we have considered assume that distressed firms access (potentially costly) contributions from existing shareholders. In practice, firms often try to sell some of their assets to pay creditors. To what extent this strategy is feasible depends in part on the covenants in the debt contract. Morellec (2001) is the first paper that carefully explores the impact of asset liquidity and covenants on optimal leverage. In his model, productivity shocks induce firms to adjust cash-generating capital stock. In the absence of financing frictions, firms prefer highly liquid assets that can be adjusted at low cost. However, flexibility has another side: it also makes it easier to sell assets in distress, potentially at fire-sale prices, to satisfy creditors. Morellec shows that asset liquidity optimally reduces leverage, because higher liquidity leads to higher frequency of asset dispositions in distress. However, when the firm includes covenants in the debt contract that prevent it from selling assets, higher flexibility increases optimal leverage.

Managerial Decisions. In the standard setup we have explored, managers act on behalf of shareholders to maximize equity value. In practice, corporate managers have their own agendas and maximize their own utility subject to the constraints imposed by the shareholders. Although central to traditional corporate finance theory, this topic has yet to be fully explored within this framework. A paper by Morellec (2004) offers one of the first attempts to incorporate managerial preferences in capital structure decisions. As in Zwiebel (1996), managers derive utility from investing and retaining control, a combination that can lead to overinvestment. In choosing debt, managers face the traditional trade-off between tax benefits and bankruptcy costs. However, they also confront a new trade-off: higher debt lessens the opportunity to invest in the future but at the same time prevents control challenges by outsiders. An important result of the calibration is that managers choose relatively low debt levels, which are easy to reconcile with low leverage observed in practice.

Carlson and Lazrak (2010) incorporate managerial preferences using a different economic mechanism. In their model, managers are risk-averse and have a linear compensation contract, with a fixed salary component and an equity stake. Their paper is based on the trade-off version of Merton (1974) with finite debt maturity, where managers choose both capital structure and an asset volatility parameter. Effectively, they choose the distribution of the final-period asset value. In relation to capital structure, managers choose higher leverage (closer to the first-best case, where managers maximize shareholders value) when the fixed component of compensation is low and interests of managers and shareholders are closer aligned.

Behavioral approach. Most dynamic contingent-claim models embed standard rationality assumptions for both decision-makers and investors. Several recent papers have started exploring the impact of deviations from agents' rationality. For example, Hackbarth (2009, 2010) build a capital structure model with investment, in which the market holds rational expectations, but the manager is both "optimistic" and "overconfident" in a sense that the manager believes the expected growth rate to be higher than the true one (optimism) and the cash flow volatility to be lower than the true one (overconfidence). Intuitively, in the presence of investment opportunities and constrained internal funding, managers end up relying on more debt issues, because debt has a lower sensitivity to a difference in opinions between the manager and the market. Managers also find it optimal to refinance more frequently.

Mergers, acquisitions, and corporate control. All the models we have discussed in detail here fix the organizational structure of the firm. An important topic that has only recently started being analyzed in dynamic contingent-claim models is the market for corporate control. For example, Lambrecht (2004) studies the timing and terms of mergers driven by economies of scale. The potential merger surplus is easy to value using the contingent claims approach. The model finds the endogenous timing of the merger as a function of firms' bargaining position, which determines the surplus. Lambrecht also shows that firms with market power optimally speed up mergers and that the equilibrium timing of hostile takeovers is inefficient. In a related paper, Lambrecht and Myers (2007) consider merger activities in a declining industry, in which managers are reluctant to liquidate optimally because they do not internalize the opportunity cost of assets in place. Using the contingent claims approach, they consider implications for a number of scenarios, such as hostile takeovers and golden parachutes.

A paper by Hackbarth and Morellec (2008) studies strategic takeover decisions, in which a pre-assigned acquirer has assets in place and a takeover option that is modeled similar to a standard real option. The trade-off that the acquirer faces is the improvement in the value of the target (for example, because of operational improvements) versus the fixed cost of the takeover. The solution of the model takes a recognizable pattern: the optimal timing of the takeover is determined by a threshold, which is a function of the ratio of cash flows of the acquirer and the target. The paper also empirically explores an interesting implication of the model concerning the evolution of betas of the acquirer. Because the acquirer has a risky takeover option before the acquisition, its beta is higher. The takeover option exercise triggers a reduction in the acquirer's beta.

Morellec and Zhdanov (2008) consider the impact of debt financing on the outcome of takeover contests. Firms with lower leverage are more likely to enter takeover contests and become an acquirer, because debt financing imposes a financing constraint on the firm and decreases the maximum price the firm is ready to pay for a target. The model predicts that the leverage of a winning bidder is lower than the industry average, consistent with empirical evidence, and that acquirers should lever up after the takeover.

Capital structure and industry competition. Capital structure can be used strategically by competitors, an issue studied originally in a static setup by, for example, Brander and Lewis (1986). Incorporating strategic competition in a dynamic setup enriches the options available to firms and the outcomes of their actions. For example, Lambrecht (2001) considers a duopolistic industry, in which both firms have pre-determined debt outstanding and thus have a real option to default and exit the industry. The remaining firm enjoys higher monopoly profits. The paper shows that financial vulnerability of the incumbent induces earlier entry of the rival and subsequent earlier default exit by the incumbent. Another approach is to study the interaction of capital structure and industry competition in a perfectly competitive industry. For example, Miao (2005) incorporates a capital structure trade-off in the model by Dixit (1989), in which firms optimally enter and exit in a perfectly competitive industry. Calibrations suggest that the stationary distribution of firms and firm's survival probabilities are influenced by capital structure decisions.

Information asymmetry. The classical study of Myers and Majluf (1984) applied the lemons problem to external financing decisions of firms. In their equilibrium, managers of good firms may choose to bypass positive NPV projects, because they do not want to be pooled with bad firms. Conditional on issuing external financing, debt financing is preferred to equity issuance, because debt is a claim with lower sensitivity and thus less dilution for knowledgeable insiders. Morellec and Schürhoff (2011) consider a real options problem where insiders are better informed about the growth option. In equilibrium, good firms separate by deciding to invest too early to prevent mimicking by bad firms. Good firms also issue debt as an additional separating tool. As a result, these considerations lead to a substantial erosion in the value of waiting.

Strebulaev, Zhu, and Zryomov (2012) consider a dynamic version of Myers and Majluf (1984), where, as in the original paper, insiders have both assets in place and a growth option and are better informed about the value of assets in place than about the growth option. The market learns the type of the firm over time even if no investment is observed. Low quality firms face the trade-off between investing now and waiting to pool with high quality firms. In equilibrium, both pooling and separation can occur, as in the static model, but comparative statics shows that the possibility of waiting changes the quantitative nature of equilibrium. For example, the probability of pooling can be substantially lower, because a low type firm finds waiting too expensive. Also, the model shows substantial delays in investment that can result in the presence of information asymmetry.

3 Discrete-Time Investment Models

Discrete-time dynamic investment models are also widely used as a backbone for understanding dynamic financing decisions.[21] As in the case of contingent claims models, discrete time investment models contain three building blocks: exogenous stochastic state variables, an objective function, and a set of endogenous state variables that can be changed via a set of control variables. The

[21] Early papers that develop this class of models include Lucas and Prescott (1971), which is a discrete-time model and Hayashi (1982), which is a deterministic continuous-time model. Although this section emphasizes the case of discrete time models, it is, of course, straightforward to write down continuous-time versions of these models (Abel and Eberly 1994; Hennessy 2004; Hennessy, Levy, and Whited 2007).

exogenous state variables are typically shocks to firm profits. As such, they can be interpreted either as demand shocks or productivity shocks. It is also possible to add further shocks, for example to production costs (Riddick and Whited 2009) or to financing costs (Jermann and Quadrini 2012). The objective function is typically to maximize shareholder value, which is the expected present value of cash flows to shareholders. Once again, it is possible have different objectives, such as maximizing CEO utility. The endogenous state variables in dynamic investment models always include the current capital stock. They can also include other variables such as labor, the stock of liquid assets, or debt. The paths of these endogenous state variables can then be altered via choice variables such as *next* period's capital, labor, liquid assets, or debt.

As we detail below, these models are more restrictive on some dimensions than are contingent claims models, and they are more flexible on others. On the one hand, contingent claims models allow for the *pricing* of debt and equity claims, whereas investment-based capital structure models typically do not. On the other hand, investment-based models allow for a richer set of financing choices. For example, whereas in contingent claims models the firm finances with either debt or equity, in investment-based models, the firm can finance with debt, cash, equity infusions, and dividend cuts. A more important advantage is that investment is chosen endogenously along with financing variables so that it is possible to study the interaction between real and financial decisions. Although it is also possible to study this interaction in real-options models, as in Morellec (2004), the class of models outlined below allows for a variety of different investment policies, as opposed to the lumpy, discrete projects that are central to real-options models.

These models can shed light on many interesting questions. For example, they have been used extensively to understand the effects of financial constraints on investment. They have also been useful for understanding corporate cash holding decisions. Finally, the most prominent example has been corporate capital structure, and in particular, the nature of the interaction between real investment and debt-equity choices.

Dynamic investment models are typically cast as infinite horizon problems. At first, the task of solving such models appears daunting, in that it appears to be necessary to choose an infinite sequence of controls. However, these problems can be greatly simplified by assuming that the exogenous state variables follow Markov processes. Loosely speaking, if a variable follows a Markov process, then its expected value given its entire previous history equals its expected value just given its previous period's value. In other words, the current value of the stochastic state variable captures its entire history. In this case, it is usually possible to characterize the model solution using just two functions. The first is the value function, which is a rule that specifies firm value as a function of the current exogenous and endogenous state variables. The second is the policy function, which is a rule that specifies next period's controls as a function of the current exogenous and endogenous state variables.

This section is structured in the same way as Section 2 in that we begin with a simple model and then add complexity. Thus, this section starts with the simplest pure investment model, which contains no meaningful financing decisions. By building up the analytical framework and highlighting the main intuition behind this type of model, we provide a solid foundation for understanding the financing models that follow. To formulate these financing models, we layer on top of this basic dynamic investment framework different financing features, with each new version of the model slightly more complex than the previous. We then relate each version of the model to specific studies in the literature.

With this general overview in mind, we now turn to the development of specific models.

3.1 Basic Model

This basic model is a partial equilibrium model cast in discrete time, with an infinite horizon.Capital is the only fixed factor of production, and risk-neutral managers, acting on the behalf of shareholders, choose the capital stock each period to maximize the value of the firm, which is the expected present value of the stream of future cash flows to (or from) shareholders. These cash flows, which we denote as $e(k_t, I_t, z_t)$ can be expressed via a sources and uses of funds identity:

$$e(k_t, I_t, z_t) \equiv \pi(k_t, z_t) - \psi(I_t, k_t) - I_t. \tag{3.1}$$

Here, k_t is the beginning-of-period capital stock; I_t is investment in capital; $\pi(k_t, z_t)$ is a profit function with $\pi_z > 0$, $\pi_k > 0$, and $\pi_{kk} \leq 0$; z_t is an exogenous shock to the profit function that is observed by the manager at time t and that follows a Markov process; and $\psi(I_t, k_t)$ is an investment adjustment cost function that is increasing in I_t and weakly decreasing in k_t, the idea being that investment is less disruptive for larger firms. It is worth noting that in contrast to the real options models described in Section 2.2, I_t can be either positive or negative. Disinvestment is costly because of the adjustment cost function, but it is not completely irreversible.

Several remarks are in order. First, for simplicity, the price of capital is normalized to unity. Second, expressing profits (not production) as a function only of capital is not restrictive in the sense that any variable factors of production, such as labor, have already been maximized out of the problem. Third, the adjustment cost function $\psi(\cdot)$ represents any reduction in profits that comes from the disruption of operations that can accompany the installation of capital. Thus, equation (3.1) says that cash flows to shareholders, $e(k_t, I_t, z_t)$, are simply profits minus all expenditures associated with investment in capital.

In this model, $e(k_t, I_t, z_t)$ can be either positive or negative. If positive, it represents distributions of internal cash flows to shareholders, and if negative, it represents infusions of cash from shareholders into the firm. Thus, in this model internal funds and external financing from shareholders are equally costly, the Modigilani-Miller theorem holds, and financing is trivial. Nonetheless, it is worth spending some time on this simple model because a great deal of the intuition behind the more complicated models that follow revolves around investment behavior.

With all of these ingredients we can express the value of the firm at time t as:

$$V_t = \max_{k_{t+j}|j=1,\ldots,\infty} \mathbb{E}_t \left[\sum_{j=0}^{\infty} \left(\frac{1}{1+r} \right)^j e(k_t, I_t, z_t) \right], \tag{3.2}$$

where $\mathbb{E}_t$ is an expectations operator conditional on information known at time t, and r is the constant risk-free interest rate.

3.1.1 First-Order Conditions

This model has no analytical solution. However, one can characterize the solution via an examination of the first-order conditions, and doing so helps lend intuition to the solution. The firm maximizes (3.2) subject to the following capital stock accounting identity:

$$k_{t+1} = (1-\delta)k_t + I_t, \tag{3.3}$$

where δ is the assumed constant rate of capital depreciation.

An informal way to obtain the first-order condition is simply to form the Lagrangian:

$$\mathbb{E}_t \left[\sum_{j=0}^{\infty} \left(\frac{1}{1+r} \right)^j \Big(e(k_t, I_t, z_t) - \chi_t (k_{t+1} - k_t(1-\delta) - I_t) \Big) \right], \tag{3.4}$$

in which χ_t is the Lagrange multiplier on the constraint (3.3). Then differentiating (3.4) with respect to I_t yields:

$$1 + \psi_I(I_t, k_t) = \chi_t. \tag{3.5}$$

Because χ_t is a Lagrange multiplier, it represents the shadow value of capital, so this first-order condition says that at an optimum the shadow value of capital equals its marginal cost, which has two components. The first is the price of capital, and the second is the marginal adjustment cost.

It is possible to obtain an intuitive expression for the shadow value of capital, χ_t, from the Euler equation, which we obtain informally by taking the first-order conditions of the problem of maximizing (3.4) with respect to k_{t+1}:

$$\mathbb{E}_t \left[\frac{1}{1+r} \Big(\pi_k(k_{t+1}, z_{t+1}) - \psi_k(I_{t+1}, k_{t+1}) + (1-\delta)\chi_{t+1} \Big) \right] = \chi_t. \tag{3.6}$$

This expression says that the shadow value of capital today equals its discounted expected value tomorrow plus any marginal profit flows that today's capital generate. To solve for χ_t, we substitute (3.6) into itself recursively and use the law of iterated expectations to obtain:

$$\chi_t = \mathbb{E}_t \left[\sum_{j=1}^{\infty} \left(\frac{1}{1+r} \right)^j (1-\delta)^{j-1} \Big(\pi_k(k_{t+j}, z_{t+j}) - \psi_k(I_{t+j}, k_{t+j}) \Big) \right]. \tag{3.7}$$

Thus, equation (3.7) shows that the shadow value of capital is the expected stream of future marginal benefits from using the capital. These benefits include both the marginal additions to profit ($\pi_k(k_t, z_t)$) and reductions in installation costs ($\psi_k(k_t, I_t)$). The quantity χ_t is usually called "marginal q" (Hayashi 1982).

It is worth noting that if one uses quadratic investment adjustment costs, then the first-order condition (3.5) turns into a linear relationship between investment and marginal q, which has been the formal basis for countless published investment regressions.

Further intuition can be gleaned from substituting (3.5) into (3.6) to obtain:

$$\mathbb{E}_t \left[\frac{1}{1+r} \Big(\pi_k(k_{t+1}, z_{t+1}) - \psi_k(k_{t+1}, I_{t+1}) + (1-\delta)(1 + \psi_I(I_{t+1}, k_{t+1})) \Big) \right] = 1 + \psi_I(I_t, k_t) \tag{3.8}$$

This optimality condition can be understood in the form of a perturbation argument. If the firm is making an optimal choice, it must be indifferent on the margin between installing a unit of capital today and waiting to install that unit of capital tomorrow. Thus, if the firm installs capital today, it incurs marginal purchasing and adjustment costs, which can be seen on the right hand side of (3.8). If the firm waits until tomorrow, we can see from examining (3.8) that it again incurs the same marginal purchasing and adjustment costs, albeit discounted by both the interest rate and the depreciation rate. However, it also incurs the opportunity cost of waiting to invest until tomorrow—the foregone marginal product of capital, which is captured by the term $\pi_k(k_{t+1}, z_{t+1}) - \psi_k(k_{t+1}, I_{t+1})$. As we explain later, this investment Euler equation has been used to study the effects of finance constraints on investment (Whited 1992; Bond and Meghir 1994).

3.1.2 Numerical Solution

Most recent studies based on variants of this simple model have used numerical model solutions to characterize both financing and investment decisions. Although it is beyond the scope of this article to provide a detailed explanation of the relevant numerical methods, it is nonetheless useful to provide an intuitive outline that emphasizes implementation rather than generality. An excellent treatment of numerical methods can be found in Miranda and Fackler (2002), and proofs of the existence of solutions to this class of models can be found in Stokey, Lucas, and Prescott (1989).

As a first step in outlining an algorithm for a numerical solution, we drop the time subscripts to simplify notation. Thus, for any generic variable x, let x_t be denoted as x, and let x_{t+1} be denoted as x'.

Next, we assume that the shock z follows a specific Markov process—a first-order autoregressive process, $AR(1)$, in logs:

$$\ln(z') = \rho \ln(z) + \varepsilon'. \tag{3.9}$$

Here, ρ is an autoregressive coefficient, and ε' is the error term in this $AR(1)$ process. It has a variance σ_ε^2. We assume that ε has a truncated normal distribution, so that z is approximately lognormally distributed. We also introduce notation for the Markov transition function for z, that is, the conditional distribution function of z' given z. We denote the transition function for z as $g(z' \mid z)$ so that the expectations operator $\mathbb{E}_t(\cdot)$ can be written as $\int(\cdot)dg(z' \mid z)$.

We can then heuristically derive the Bellman equation corresponding to (3.2) by splitting (3.2) into two parts: the term corresponding to time t (the first period's net profit) and the rest of (3.2). Substituting the capital stock accounting identity (3.3) into the sources and uses of funds identity (3.1), we obtain:

$$V(k,z) = \max_{k'} \left\{ \pi(k,z) - \psi(k' - (1-\delta)k, k) - (k' - (1-\delta)k) + \frac{1}{1+r} \int V(k', z') dg(z' \mid z) \right\}. \tag{3.10}$$

The first three terms of (3.10) constitute the current period flow to shareholders: profits less investment less adjustment costs. The last term is called the continuation value, that is, the expectation of next period's firm value, given that optimal policies will be chosen in the future.

The goal of a numerical solution is to obtain a mapping from (k, z) to $V(k, z)$. The simplest, possibly the slowest, but often the most robust, method for obtaining a solution is called discrete state value function iteration. Here, we describe the basic algorithm. It starts with a choice of grids for the state variables (k, z). For example, one might stipulate that the capital stock k can only take three values, (49, 50, 51), and that the shock can only take on two values, (0.9, 1.1). In practice, these grids are much larger. Further, it is important to center the capital stock grid at the steady-state level of the capital stock, which is usually a good approximation to a model solution corresponding to the mean level of the shock, z. This level is roughly the point at which the marginal product of capital corresponding to $\ln(z) = 0$ (approximately $z = 1$) equals the user cost of capital, $r+\delta$. It is also important to have the densest part of the grid correspond to the parts of the production function with the most curvature. The reason is that numerical solutions are usually linear approximations to nonlinear functions, and line segments are good approximations to functions with a great deal of curvature only over short intervals. One way to accomplish this design is to let the capital stock grid be a multiplicative sequence so that $k^{i+1} = k^i/(1-\delta)$, in which k^{i+1} and k_i are adjacent values in the grid.

We need one more ingredient before we can explain the algorithm: the transition function for z. If z follows an $AR(1)$ in logs as in (3.9), then one can use a variety of methods (Tauchen 1986), Tauchen and Hussey (1991) to construct a discrete state Markov chain that approximates (3.9). One then ends up with a grid of possible z values and a transition function in the form of a Markov transition matrix.

With all of these ingredients we can now describe value function iteration. As a first step, we characterize the form of the value function. Let n_z be the number of points in the shock grid and n_k be the number of points in the capital stock grid. The value function is then an $n_z \times n_k$ array of numbers, with each number corresponding to the firm value that is associated with a specific (k, z) pair.

The simplest form of this algorithm then proceeds as follows: First, guess a value function $V(k', z')$; for example, an $(n_z \times n_k)$ array of zeros suffices. Second, take the expectation of the value function by multiplying it by the Markov transition matrix. Third, using this expected value function, evaluate the right-hand side of (3.10) for all possible (k, k', z) triples. Note that the profit flow in (3.10) needs to be evaluated for every (k, k', z) triple. We denote the result as $\tilde{V}(k, k', z)$. Fourth, compute $\max_{k'} \tilde{V}(k, k', z)$. The result is a new $V(k, z)$. Fifth, use this result as a new candidate $V(k', z')$, and repeat until the value function converges. Once one has obtained a value function, one then picks out the optimal policies that produce the value function. This mapping of states, (k, z), to policies, k', is called a policy function, which we denote by $h(k, z)$.

Once one has obtained the policy function, one can generate a time series of optimal capital stocks. As a first step, one simulates a series of z shocks. Then one uses the optimal policy function to trace out the optimal capital stock path. For example, suppose the shock z can take on only two values and that the Markov transition matrix is:

$$\begin{pmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{pmatrix}.$$

Thus, if one is in the first state, the probability of remaining in the first state is 0.6, the probability of transitioning to the second state is 0.4, and so on. To simulate the path of shocks, one generates a random uniform variable. Suppose the shock is currently in the first state. If the random uniform number is greater than 0.6, the shock moves to the second state, and otherwise the shock stays in the first state. This algorithm is straightforward to extend to multiple shocks.

3.1.3 Model Intuition

To illustrate the model intuition, we examine a special case of this basic model in which $\pi(k, z) = zk^{\theta}, 0 < \theta < 1$. We consider three forms of the adjustment cost function: no adjustment costs, convex adjustment costs, and fixed adjustment costs.

No Adjustment Costs

With no adjustment costs, the model has an analytical solution for the policy function, $k' = h(k, z)$. To derive the policy function, we note that without adjustment costs the Euler equation given by (3.8) simplifies to:

$$\int \theta z' k'^{\theta-1} g(z' \mid z) = r + \delta \tag{3.11}$$

Given the assumption that the shock z follows (3.9), we use the properties of a lognormal distribution to write (3.11) as:

$$\exp\left(\rho \ln(z) + \frac{1}{2}\sigma_z^2\right)\theta k'^{\theta-1} = r + \delta \tag{3.12}$$

The exponential term on the left-hand side of (3.12) is the expected value of z', given z, and given that z has a lognormal distribution. Thus, tomorrow's optimal capital stock, k', is set so that the expected marginal product of capital (the left-hand side of (3.12)) equals the user cost of capital (the right-hand side). For the special case of no adjustment costs, the optimal choice of k' only depends on the shock, z, and does not depend on k. This investment model without adjustment costs can thus be broken down into a series of one-period problems. Each period, the firm observes z and then chooses optimal k', *regardless* of the current value of the capital stock. Later, we will see that the introduction of investment or financing frictions introduces dependence between periods and thus real dynamics.

An examination of (3.12) reveals five properties of these models that carry through to the more complex models we examine later. First, capital (and thus investment) are naturally increasing in the productivity/demand shock, z. Second, the average rate of investment is largely determined by δ because capital depreciates by δ every period and because the expected value of z is approximately 1. Third, the smaller θ, the less variable is investment. Intuitively, the more concave the production function, the less the production function shifts in response to the multiplicative shock, z. In economic terms, if the firm has sharply decreasing returns, then technology shocks have only a modest effect on the marginal profit of capital, so the firm responds less strongly to these shocks. Thus, the optimal capital stock moves less than it would in the case of a less concave production function. Fourth, the more variable z, the more variable investment. Fifth, if ρ equals zero, the firm never changes its optimal capital stock because it knows that any productivity increases will be fleeting. As ρ increases, the firm responds more strongly to shocks because it knows that the marginal product of capital is likely to last. This last property would be impossible to obtain from a static model because in such a model the notion of serial correlation is nonexistent.

Convex and Fixed Adjustment Costs

Next, we consider the case of an adjustment cost function that encompasses both fixed and convex adjustment costs, so that $\psi(I,k) = \psi_0 I^2 \big/ 2k + \psi_1 k \mathbb{I}_{I \neq 0}$. The first term is the convex cost term, and the second term is the fixed cost term, which does not depend on the amount of investment being undertaken. Note that adjustment costs are incurred for both positive and negative investments. In this case, the model has no analytical solution, so we explain the intuition behind the model for a special case in which $\psi_0 = 0.01$, $\psi_1 = 0$, $r = 0.04$, $\delta = 0.15$, the production function curvature, θ is 0.7, the serial correlation, ρ, of the shock process (3.9) is 0.7, and the standard deviation of the error term in (3.9), σ_ε, equals 0.15. We set the adjustment cost parameters at low values for expositional purposes, so that the intuition behind the rest of the model is transparent. The rest of the parameter choices are standard and close to those used in papers such as Gomes (2001) and Hennessy and Whited (2005, 2007).

Figure 10 illustrates the intuition behind the role of adjustment costs by plotting the policy function for the ratio of investment to the capital stock, $(h(k,z) - (1-\delta)k)/k$, for three models: one with no adjustment costs, one with smooth adjustment costs, and another with fixed adjustment costs. The policy function is evaluated at the steady-state capital stock and is expressed as a function of the log shock, $\ln(z)$. The magnitudes can be interpreted approximately as the percent increase or decrease in profits relative to k^θ. We see that investment rises sharply when there are no adjustment costs, with a two-standard deviation shock producing a rate of investment of over 100%. This behavior is damped substantially in the presence of smooth adjustment costs. Fixed costs produce a large range of inaction for shocks ranging from four standard deviations below the mean to almost two standard deviations above the mean. At that point the firm optimally invests a great deal.

Figure 11 fleshes out this intuition by depicting comparative statics exercises in which we show how various moments of the simulated data change with respect to the parameters. We construct each panel of Figure 11 as follows. We start with the baseline parameterization, and then allow one parameter to vary over a range of 20 values, which is shown on the horizontal axis of each figure.

For each of these 20 parameter values, we solve the model and generate simulated data. We then calculate the moments of interest. The first moment is average investment, which is defined as I/k. The next is the financing surplus/deficit (as a fraction of the capital stock), which is defined as $\left(\pi(k,z) - \psi(k' - (1-\delta)k, k) - (k' - (1-\delta)k)\right)\Big/ k$ and thus equals cash inflows minus cash outflows. External financing equals the absolute value of the surplus/deficit variable, when it is negative, and zero otherwise. It too is scaled by the capital stock.

The first panel shows that production function curvature has a strong impact on investment and financing policies. In particular, when θ is small and the firm has sharply decreasing returns to scale, it does not respond strongly to shocks, and its optimal investment rarely requires external financing. Indeed, it invests *exactly* enough to replace depreciated capital 45% of the time, and this amount of investment is easily met by internally generated funds. This behavior occurs because the effect of the shock, z dies out fairly rapidly: after t periods, the effect is ρ^t. Therefore, with θ small, the marginal product of capital does not move much with z, so the firm does not respond strongly, if at all, to small movements in z. In contrast, when the firm has nearly constant returns to scale, its investment equals capital depreciation only 3% of the time and is highly variable. This behavior is the result of two model features. First, because the production function is nearly linear, movements in z change expected marginal profitability a great deal. Second, the concavity of the profit function means that the multiplicative shocks have an asymmetric effect. Positive shocks entail large shifts in the marginal product of capital and thus in the optimal capital stock, and negative shocks entail smaller shifts. Thus, we see many instances in which the firm finds it optimal to make big investments, which require external financing. Therefore, for high values of θ, on average the firm runs a financing deficit and uses a great deal of external financing. Mitigating but not reversing these effects is the fact that a higher level of θ implies higher average profits, which in turn imply a lower need for external financing.

The second panel shows that the depreciation rate has a more modest effect on investment and financing. Average investment is slightly higher than the depreciation rate, and average external financing rises with investment, but on average the firm generates more internal funds than are needed for investment. The reason for the muted response is that although changes in the depreciation rate do affect the value of capital and the average rate of investment, these changes do not affect the *sensitivity* of the marginal product of capital to the shock z.

The third panel shows that the standard deviation of the shock process also matters a great deal for average investment and financing in much the same way as does the curvature of the profit function. In this case the intuition is similar. With highly volatile shocks, the firm occasionally receives a great investment opportunity and requires outside financing, but the concavity of the profit function implies that the same effect does not occur for negative shocks, which entail no or negative investment. Thus, investment and external financing increase with the standard deviation, σ_ε, and the financing deficit decreases.

The fourth panel appears to show that an increase in the serial correlation of the shock process, ρ, has a similar but somewhat more muted effect on average investment and financing. The intuition is similar. With highly serially correlated shocks, when the firm receives a positive shock, it is more likely to make a large investment than when it has *i.i.d* shocks because it expects the marginal product of capital to last. Thus when ρ increases, the probability of large investments increases, as does the need for financing.

The fifth and sixth panels show that convex and fixed investment adjustment costs have opposite effects on investment and financing. When convex costs increase, the optimal policy of the firm is to smooth investment over time, and for very high costs, the firm simply invests to replace depreciated capital so that the average rate of investment equals the depreciation rate. In this case the firm never needs outside financing and always runs a financing surplus. In contrast, in the case of high fixed costs, the firm invests in a lumpy fashion. Thus, the distribution of investment becomes highly skewed, which mechanically raises the mean of investment. Also, on average the firm needs to raise finance because the large bursts of investment are larger than internally generated funds.

Three key takeaways emerge from this exercise. First, because of a standard sources-and-uses-of-funds identity, the productivity shock has a profound effect on financing decisions via its effect on investment decisions. Second, technology and uncertainty matter greatly for optimal financial policies because they dictate the variability and lumpiness of investment policy. Thus, understanding basic investment dynamics is essential for understanding financing decisions. Third, the level of external flow financing is much higher than what one observes in the data because the marginal cost of external financing is the same as that of internal financing. The average amount of financing raised by firms in Compustat is typically less than 10% of assets in any given year, whereas in the simulated data, it is around 20% of assets.

3.2 Costly External Finance

We now introduce costly external financing into this basic model. The simplest formulation comes from Gomes (2001), in which external finance takes the form of equity injections from current shareholders. These injections carry a fixed and proportional cost, and are thus a more expensive source of funds than internally generated cash flows. This structure follows Altinkilic and Hansen (2000), which estimates that underwriting fees are characterized by both fixed and variable components. The firm does not use debt financing, and the firm cannot retain earnings. If cash inflows exceed optimally chosen cash outflows, the firm must pay the entire excess out to shareholders. On the other hand, if outflows exceed inflows, then to fill the gap, the firm pays transactions costs. Define:

$$e(k, k', z) \equiv \pi(k, z) - \psi(k' - (1-\delta)k, k) - (k' - (1-\delta)k), \tag{3.13}$$

that is, cash inflows ($\pi(k, z)$) minus cash outflows ($-\psi(k' - (1-\delta)k, k) - (k' - (1-\delta)k)$). First, note that (3.13) is identical to (3.1), except with simpler notation. It is also important to note that this sources and uses of funds identity implies that if one chooses capital, then one also implicitly chooses $e(\cdot)$. It is impossible to choose both separately. Next, as in Gomes (2001), we define the cost of external finance as:

$$\eta(e(k, k', z)) \equiv \left(\eta_0 + \eta_1 e(k, k', z)\right) \mathbb{I}_{e<0}, \tag{3.14}$$

in which $\mathbb{I}$ is an indicator function. The cost of external finance is assumed to be zero if the firm is distributing funds to shareholder, but it is strictly positive if the firm is receiving injections of funds from shareholders. In particular, the function contains a fixed component, η_0, as well as a linear component, η_1.

With costly external finance the Bellman equation for this model is:

$$V(k,z) = \max_{k'} \left\{ e(k,k',z) + \eta(e(k,k',z)) + \frac{1}{1+r} \int V(k',z') dg(z' \mid z) \right\} \tag{3.15}$$

The model in Gomes (2001) is substantially richer than this simple partial equilibrium framework. His model excludes capital adjustment costs but includes a fixed cost of production, and he explicitly models both labor and capital choices. More importantly, his model is an industry equilibrium model. A representative consumer consumes the firm's output and supplies labor to the firm. Then the real wage adjusts so that the supply of labor equals the demand for labor. Firms enter an industry when they get a shock that is sufficiently high for them to break even. Firms exit when the proceeds from selling capital are greater than the value of the firm as a going concern. The result of this entry and exit is a steady state distribution of firms.

In this setting, it is possible to characterize four types of firms: those that exit, those that do not need external finance, those that both need and choose to obtain external finance, and those that would obtain external finance if it were costless but that choose not to do so when it is costly. In general, firms with extremely low productivity exit. For those firms that do not exit, their financing status depends on both their size and productivity, with large, lower productivity firms not needing finance, intermediate sized, intermediate productivity firms being constrained, and small, extremely productive firms getting finance.

Although the model in Gomes (2001) is rich enough to explore a wide variety of questions, one particularly important part of the paper is an attempt to answer the question of whether the sensitivity of investment to cash flow is a good measure of finance constraints. This question is from Fazzari, Hubbard, and Petersen (1988), who argue that the investment of constrained firms should be highly correlated with movements in internal funds. They then test this proposition by running panel regressions of the ratio of investment to capital on the ratio of cash flow to capital and Tobin's q, which is the ratio of the market value of the capital stock to the replacement value.

In our notation, this regression can be expressed as:

$$\frac{I}{k} = b_0 + b_1 \frac{V(k, k', z)}{k} + b_2 \frac{zk^\theta}{k} + u, \tag{3.16}$$

in which b_0, b_1, and b_2 are regression coefficients and u is a regression disturbance. Fazzari, Hubbard, and Petersen (1988) find that the sensitivity of investment to cash flow (b_2) is highest for groups of firms they categorize as constrained.

The model in Gomes (2001) is ideal for evaluating whether investment regressions are useful for detecting finance constraints for two reasons. It is possible to run this precise regression on simulated data, and it is possible to know exactly which simulated firms are constrained and which are not. The main result in Gomes (2001) is that investment-cash flow sensitivity is neither sufficient nor necessary for the existence of finance constraints.[22] The intuition is that in a dynamic model with entry, exit, and costly external finance, the first-order conditions of the model are only weakly approximated by a regression of investment on Tobin's q and cash flow. If one could estimate the first-order conditions of this model, one might be able to identify constrained firms, but using an approximation falls short in this dimension.

We now depart from the analysis in Gomes (2001) to show that this point can also be seen in a slightly different way in the simpler framework here. We do so by examining what happens to investment-cash flow sensitivity when we change the parameters that govern the cost of external finance, η_0 and η_1. To this end we initially set $\eta_0 = 0.08$ and $\eta_1 = 0.028$, as in Gomes (2001).

The top two panels of Figure 12 presents the results from this comparative statics exercise. Each panel is constructed exactly as for Figure 11, with the exception of the top two panels. In these panels, which explore the costs of external finance, when we vary η_0, we set $\eta_1 = 0$, and vice versa. On the vertical axis of each panel is the regression coefficient b_2.

Several results are of note. First, in both panels, even when the cost of external finance is zero, the cash flow coefficient is positive. The reason is that with decreasing returns to scale, Tobin's q is not a sufficient statistic for investment (Hayashi 1982). Thus, as pointed out in Cooper and Ejarque (2003), because cash flow contains incremental information about investment opportunities,

[22] See also Alti (2003), which examines related partial equilibrium model with learning.

it enters into the regression. Second, the cash flow coefficient is decreasing in the cost of external finance, not increasing, as predicted by Fazzari, Hubbard, and Petersen (1988). The reason, as pointed out in Moyen (2004) and Hennessy and Whited (2007), is that when the firm cannot freely access external finance, it cannot respond strongly to profit shocks by investing. Thus, when the cost of external finance rises, investment becomes less highly correlated with the information about investment opportunities in cash flow, and therefore with cash flow itself. Note that this intuition is dramatically different from the simple, and static, argument in Fazzari, Hubbard, and Petersen (1988) that revolves around increasing the firm's liquid resources in a one-time, lump-sum fashion. The realistic and nuanced intuition from a dynamic model is thus difficult to obtain from a static model. Third, the coefficients from the simulated data are approximately 10 times as large as the coefficients one finds in real data. This result, which is also in Gomes (2001), is to be expected in this simple model for two reasons. It contains minimal investment frictions, and with only one choice variable, no other choices break the strong correlation between profit shocks, investment, and cash flow.

The next six panels of Figure 12 make a different point. Financial frictions are by no means the only firm characteristics that matter for investment-cash flow sensitivity. As in Figure 11, we plot the cash flow coefficient as a function of the six model parameters that describe technology and uncertainty. One guiding piece of intuition informs all of these comparative statics exercises: any parameter that makes investment respond more strongly to z increases investment-cash flow sensitivity, and vice versa.

The third panel shows that cash flow sensitivity increases as the production function becomes closer to linear, i.e., as θ approaches one. A nearly linear production function implies large swings in the optimal capital stock when z fluctuates. Therefore, investment responds more strongly to z, which leads to increased cash flow sensitivity. The fourth panel shows that changes in the depreciation rate have little to do with the response of investment to shocks, because the depreciation rate helps determine the average level of investment—not investment dynamics.

The fifth and sixth panels show how investment-cash flow sensitivity changes with the parameters that govern the shock process (3.9). First, we consider the standard deviation of the profit shock. With a low variance shock, investment varies little with movement in the shock, and with a high variance shock investment responds strongly to z. Recall that this pattern stems from the inherently temporary nature of the z shocks, so that the firm optimally ignores small shocks. Therefore, cash flow sensitivity increases with the shock standard deviation, although most of this effect occurs at low shock standard deviations. Now we consider the shock serial correlation. Equation (3.12) implies that investment responds to z more when the shock process is more highly serially correlated, so cash flow sensitivity increases with ρ.

Finally, the seventh and eights panels show that increasing either smooth or fixed capital stock adjustment frictions almost always lowers investment cash flow sensitivity by making investment insensitive to shocks. In addition, as fixed costs become more important, investment responds negatively to movements in cash flow. This result comes from the extreme lumpiness observed in investment as fixed costs rise. At points where the firm invests, investment opportunities plummet because of decreasing returns. Because in the regression (3.16) both Tobin's q and cash flow pick up movements in investment opportunities, investment-cash flow sensitivity is negative.

Two main takeaways emerge from this figure. First, financial frictions actually lower rather than increase investment cash flow sensitivity. Second, financial frictions are but one of many different factors that affect this correlation, so that it is difficult to attribute differential investment-cash flow sensitivity to differences in financial frictions.

3.3 Cash

We now add another layer to the model by introducing a new source of financing, the stock of liquid assets, which we denote as p, with $p \geq 0$. We assume that cash balances are held as a one period discount bond. To model the cost of holding cash, we also add corporate taxation, so that both profits and interest on cash balances are taxed at a rate τ. We also assume that depreciation expense it tax deductible. In this case the sources and uses of funds identity becomes:

$$
\begin{aligned}
e(k,k',p,p',z) \equiv\ & (1-\tau)\pi(k,z) + \tau\delta k \\
& - \psi(k'-(1-\delta)k,k) - (k'-(1-\delta)k) \\
& - p'/(1+r(1-\tau)) + p. \qquad (3.17)
\end{aligned}
$$

The first two terms, $((1-\tau)\pi(k,z)+\tau\delta k)$, are simply after tax cash flows, with the term $\tau\delta k$ representing the extra cash flow the firm receives because of the depreciation deduction. The next two terms, $(-\psi(k'-(1-\delta)k,k)-(k'-(1-\delta)k))$, represent investment in physical assets, and the last two terms, $(-p'/(1+r(1-\tau))+p)$, represent investment in liquid assets.

The Bellman equation looks much as it did in the previous section:

$$
V(k,p,z) = \max_{k',p'} \left\{ e(k,k',p,p',z) + \eta(e(k,k',p,p',z)) + \frac{1}{1+r}\int V(k',p',z')dg(z' \mid z) \right\}. \qquad (3.18)
$$

In comparing this model of cash holding with a simple investment model, the main differences are in the sources and uses of funds identity and in the existence of costly external finance. This model is a slightly simpler version of those studied in Riddick and Whited (2009), and it is closely related to the continuous time model studied in Bolton, Chen, and Wang (2011). The numerical solution procedure is similar to that described in Section 3.1.2, except that one must optimize over a pair of choice variables, (k',p'), instead of one, k'. Therefore, if n_p is the number of grid points for the cash state variable, the value and policy functions have dimensions $(n_z \times n_k \times n_p)$, and one must optimize over $(n_k \times n_p)$ pairs of state variables. The increased computational burden from having just one extra state variable is nontrivial, and models quickly become intractable when the number of state variables exceeds five or six.

Although this model also has no analytical solution, it is instructive to write down the first-order conditions. The first-order conditions for optimal cash balances can be obtained by differentiating the Bellman equation (3.18):

$$
1+\eta_1 \mathbb{I}_{e<0} = \frac{1+r\,(1-\tau)}{1+r}\int \left(V_p\big(k',p',z'\big)+\zeta'\right) dg\big(z',z\big)\,, \qquad (3.19)
$$

in which ζ is a Lagrange multiplier on the constraint that the state variable p be positive. The right-hand side of (3.19) represents the expected discounted shadow value of cash balances, and the left-hand side represents the marginal cost of external equity finance. Thus at an optimum the firm equates the shadow value of cash with the opportunity cost of using cash, which is the use of external finance.

Equation (3.19) highlights several pieces of intuition behind this model. First, cash derives value because it is an alternative to costly external finance. We call this benefit the "financial flexibility" benefit. Indeed, in this model the firm optimally holds no cash if external finance is free. Second, the term $1-\tau$ on the right-hand side of (3.19) indicates that because interest is taxed, holding cash is costly for the firm. Thus, an optimal interior cash policy balances the flexibility benefit with the tax cost. In general, this tax cost can represent a variety of costs from holding too much cash, such as agency costs.

To glean further intuition, we again use a numerical solution to the model to examine the policy function and to conduct comparative statics exercises with respect to the various model parameters. For these exercises, we use the same parameterization as the one used to construct Figure 11, except that we set $\eta_0 = 0$, $\eta_1 = 0.07$, and $\tau = 0.2$.

Figure 13 plots optimal investment, cash, and distributions to shareholders as a function of the productivity shock, z. All three variables are scaled by the capital stock, and as in Figure 10, the policy functions are evaluated at the steady state capital stock. As in Figure 10 optimal investment rises with the productivity shock, and optimal cash balances fall with the productivity shock. Thus, we see that financial and physical assets are substitutes. For high levels of the productivity shock, the firm draws its cash balances down to zero because capital becomes highly productive. For high values of z, distributions to shareholders become negative, that is, the firm raises external equity finance. The effect of costly external finance is also apparent in the muted response of investment to the productivity shock once optimal cash balances are zero and the firm needs to fund investment with external finance.

Next we turn to comparative statics exercises. We examine how the ratio of cash to book assets, $p/(p+k)$ responds to each model parameter. Once again, when we perform a comparative statics exercise for η_0, we set $\eta_1 = 0$. Figure 14 presents these exercises. The central intuition behind the results in this figure is that any model feature that raises the probability of needing external finance in the future also raises the shadow value of cash and thus optimal cash levels. Thus, the intuition from the simplest investment model plays a central role in understanding cash accumulation because optimal investment dynamics determine the probability of needing outside finance.

Of course, a necessary condition for holding cash is that external finance be costly. This feature of the model can be seen in the top two panels of Figure 14. In both cases, optimal cash holding is zero when external finance is costless. However, cash increases sharply for low costs of external finance but then flattens out. In other words, firms do not change their cash balances sharply if they face more costly external finance.

The next six panels show that other features of the firm do matter a great deal for average cash balances. The second row of panels shows that profit function curvature, θ, has a nonmonotonic effect on cash balances. Two opposing forces are at work. As the production function becomes more linear, the firm tends to invest in a lumpier fashion. These large investments are likely to require costly outside finance, so the firm holds precautionary cash balances. On the other hand, as the production function becomes more linear, the firm becomes more profitable, and it has more internally generated cash flows to finance investment. The first force is stronger when θ is low, but the second force is stronger when θ is high. The second row of panels also shows that optimal cash balances increase with the depreciation rate. The intuition is simple. Firms whose capital depreciates rapidly are more likely to need external financing. Further, a higher depreciation rate makes capital less valuable relative to the other asset held by the firm—cash.

The third row of panels shows that the variance and serial correlation of the profit shock process have strong effects on cash balances, once again, because of investment dynamics. Both high variance and highly serially correlated shocks increase the probability of needing external funds for investment, so average cash balances are increasing in both of these parameters. Not surprisingly, final row of panels again shows that investment dynamics matter. When the firm has high smooth adjustment costs, it invests only to replace depreciated capital and thus never expects to need outside funding. It optimally holds no cash. When the firm has high fixed adjustment costs, it invests in a lumpy investment, expects to need outside financing, and holds a great deal of cash.

This model has some limitations. First, because the firm has no debt, it never suffers financial distress and thus never needs to hold cash to avoid financial distress. We relax this restriction below. Second, in reality a great deal of the demand for cash derives from the role of cash as a form of working capital. For example, it is useful to have cash on hand in order to pay suppliers if sales revenues fall short of expectations. This simple model abstracts from this source of liquidity demand. Third, there is no strategic demand for cash that might arise if a firm wants to use cash to purchase a patented technology. These last two sources of the demand for cash are an interesting avenue for future quantitative research.

3.4 Risk-Free Debt

The next layer we add to this basic structure is risk-free debt as in Hennessy and Whited (2005) and DeAngelo, DeAngelo, and Whited (2011). We drop the state variable for cash, p, and we denote the stock of debt as b, where debt is held as a one-period discount bond. In these two models, the state variable b takes both positive and negative values, with a positive value denoting debt, and a negative value denoting cash. One can therefore interpret cash as negative debt—a model feature that we relax later. In order for the debt to be risk free, the lender requires that the firm be able to repay the debt by selling capital or by using its current after-tax cash flows, even in the worst state of the world. For illustrative purposes, we assume that only a fraction of the capital stock, $s = 0.5$ can be liquidated to repay capital. The collateral constraint can thus be expressed as:

$$b' \leq (1-\tau)\underline{z}k'^{\theta} + \tau\delta k' + sk', \tag{3.20}$$

in which $\underline{z}$ is the lowest possible value that the shock z can attain. In this model, the sources and uses of funds identity looks almost identical to its counterpart in the cash model from section 3.3:

$$e(k,k',b,b',z) \equiv (1-\tau)\pi(k,z) - \psi(k'-(1-\delta)k,k) - (k'-(1-\delta)k) + \frac{b'}{1+r(1-\tau)} - b. \tag{3.21}$$

The only substantive difference is one of interpretation. In the cash model from Section 3.3, $e(k,k',p,p',z) < 0$ means that the firm is raising generic external finance: there is no distinction between debt and equity. Here, $e(k,k',p,p',z) < 0$ means specifically that external *equity* finance can only be attained at a premium. In this case, the Bellman equation becomes:

$$V(k,b,z) = \max_{k',b'} \left\{ e(k,k',b,b',z) + \eta(e(k,k',b,b',z)) + \frac{1}{1+r} \int V(k',b',z')dg(z' \mid z) \right\}, \tag{3.22}$$

subject to the constraint given by (3.20):

This simple model is only slightly more complex than the model from Section 3.3, with the main additions being a collateral constraint and a financing state variable, b, that can take both positive and negative values.

The model closely resembles that in DeAngelo, DeAngelo, and Whited (2011). One important difference is that in DeAngelo, DeAngelo, and Whited (2011), debt is constrained to be less than a fixed fraction of the steady state capital stock given by (3.12). Our model also resembles that in Hennessy and Whited (2005), except that the latter is substantially richer in detail. In particular, the model in Hennessy and Whited (2005) contains both corporate and personal taxation, as well as a strictly convex corporate tax schedule. As is the case here, debt can be collateralized by profits, but the rest of the collateral constraint is slightly more elaborate. Instead of assuming that a fixed fraction of the capital stock can be used as collateral, Hennessy and Whited (2005) employ the closely related assumption that the firm can sell up to its entire capital stock to repay debt, but that the price is lower than 1. In other words, "fire sales" of capital occur at a discounted price occur if the firm does not have enough internally generated cash flow to repay its one-period debt. Although the model features risk-free debt, the fire sales capture the notion of financial distress. The model also does not contain capital adjustment costs, a feature that simplifies the fire sales in the model.

Nonetheless, the main intuition behind our simple model is largely the same as that behind the models in Hennessy and Whited (2005) and DeAngelo, DeAngelo, and Whited (2011), and the discussion that follows draws from both of these papers. To begin the development of this intuition, we examine the first-order condition for optimal debt/cash accumulation, which we obtain by differentiating (3.22) with respect to b'. As in Section 3.3, let $\eta_0 = 0$ and $\eta_1 > 0$. Also, let ζ be the Lagrange multiplier on the constraint given by (3.20), then the first-order condition is:

$$1 + \eta_1 \mathbb{I}_{e<0} = -\frac{1 + r\,(1-\tau)}{1+r} \int \left(V_b\big(k', b', z'\big) + \zeta'\right) dg\big(z', z\big)\,, \tag{3.23}$$

This first-order condition conveys much of the same intuition as the first-order condition for optimal cash holding in Section 3.3. As seen on the left-hand side of (3.23), debt derives value because external equity financing is costly. As seen on the right-hand side, debt also derives value because of the tax benefit of debt.

Further intuition can be gleaned from using the envelope condition,[23] which we can use to rewrite (3.23) as:

$$1 + \eta_1 \mathbb{I}_{e<0} + \zeta = \frac{1}{1+r} \int \left(1 + r\,(1-\tau)\right)\left(1 + \eta_1 \mathbb{I}_{e'<0} + \zeta'\right) dg\big(z', z\big)\,. \tag{3.24}$$

To interpret this condition, we follow Hennessy and Whited (2005) by taking the current state (k, b, z) as given, and then imagining that the optimal investment decision has already been made. In this case, all of the terms in (3.21) are constant, except for the one involving b'. Therefore, a dollar increase in debt either increases distributions or decreases equity issuances. Further, (3.21) implies a dollar increase in debt today leads to more debt repayment in the future. These two simple points are crucial to understanding the intuition behind the model.

The left-hand side of (3.24) represents the marginal benefit of using an extra unit of debt. To see this point, note that if the firm's current state and optimal investment policy imply that it has

[23] At an optimum, the derivative with respect to the first period profit flow must equal the derivative with respect to the value function.

a financing deficit (i.e. $e < 0$), then an extra dollar of debt financing keeps the firm from having to use costly external equity financing today, and the benefit of this extra dollar of debt is $1 + \eta_1$. On the other hand, if the firm has a financing surplus, then using a dollar of debt financing means that the firm can distribute an extra dollar to shareholders. Because we are not modeling distribution taxes, the benefit of debt is simply 1.

The right-hand side of (3.24) represents the marginal cost of debt financing, which can be seen to be the expected principal and interest on debt that must be repaid tomorrow. Two extra terms add texture to this simple intuition. First, the term $\eta_1 \mathbb{I}_{e'<0}$ implies that the marginal cost of debt is higher when the firm expects to have a financing deficit next period, ($e' < 0$). In other words, raising an extra dollar of debt today implies debt repayment tomorrow and therefore a higher likelihood of needing external equity financing tomorrow. Second, the presence of the Lagrange multiplier in (3.24) shows that the marginal cost of debt is also higher when the firm expects to bump up against its collateral constraint next period, that is, exhaust its debt capacity. As explained in DeAngelo, DeAngelo, and Whited (2011), the intuition is that choosing a high level of debt today implies lowers financial flexibility in the future, so the cost of borrowing today includes the value lost when a firm fails to preserve the option to borrow later. This marginal cost schedule is thus increasing in b', because raising b' increases the likelihood of resorting to positive equity issuance next period and increases the option value of debt capacity preservation.

Further intuition can be gleaned from plotting the first-order condition (3.24). Figure 15, adapted from Hennessy and Whited (2005), does so by plotting the marginal cost and marginal benefit of debt. Along the horizontal axis is b'. We do not distinguish between $b' > 0$ and $b' < 0$, although points to the left are more likely to represent cash and points to the right are more likely to represent debt. The upward sloping line is the marginal cost of debt, and the two horizontal lines represent the two possible values of the marginal benefit of debt, with the solid portion representing the entire marginal benefit schedule. Note that the solid portion jumps down at the point $b'_{e=0}$, which represents the (possibly suboptimal) level of b' at which the financing deficit is equal to zero. For points to the left of $b'_{e=0}$, the firm is not generating enough resources via debt and profits to fund its optimal investment policy. It therefore must resort to external equity issuance, and the marginal benefit of debt is therefore $1+\eta_1$. At the point $b'_{e=0}$, the marginal benefit schedule jumps down to 1 because for levels of debt beyond this point an extra unit of debt serves to increase distributions to shareholders.

Optimal debt policy is then given by the point at which the marginal costs and benefits of debt are equal. Figure 16 depicts a situation in which the firm has abundant sources of funds or few optimal uses of funds, so that the level of debt that sets the financing surplus equal to zero is low. The point $b'_{e=0}$ can be low for two reasons. First, the firm might have low (or even negative) debt, so that it requires few funds to repay debt in the future. Second, the firm might receive a low productivity shock, which implies a low marginal product of capital, which in turn implies that it is preferable to distribute profits to shareholders than invest them in the firm. In this case, the optimal level of debt is L, and shareholder distributions are then the difference between L and $b'_{e=0}$.

Figure 16 is an analogous figure that depicts optimal debt policy when the firm has limited sources of funds but many optimal uses. In this case, $b'_{e=0}$ is high either because the firm has high debt going into the period, which requires repayment, or because the firm receives a high productivity shock so that it is optimal to invest profits rather than distribute them. The firm's optimal debt level is given by H, and equity issuance is the difference between $b'_{e=0}$ and H.

These two figures highlight two important pieces of intuition. First, leverage is likely to be positively serially correlated; that is, if one were to run an autoregression on leverage, one ought to find a high coefficient on lagged leverage. Higher lagged debt causes the firm to occupy the high portion of the marginal benefit schedule over a longer stretch. With higher lagged debt, more debt must be issued this period before the marginal unit of debt serves to increase distributions rather than to replace external equity. Second, highly liquid firms are likely to be debt conservative. In this case debt issuance serves to finance higher distributions to shareholders, rather than replacing costly external equity. Since high liquidity firms occupy the lower portion of the marginal benefit schedule, debt issuance is less attractive.

To elaborate on the model intuition, we again plot the model policy function, which expresses optimal investment, debt (or cash, if negative), and distributions as a function of the log productivity shock. As before, the policy function is evaluated at the steady state capital stock, and all variables are scaled by the capital stock. As seen in Figure 17, investment rises with the productivity shock. Interestingly, for low levels of the productivity shock, investment is financed by dissaving (when the state variable $b < 0$) or by increases in debt (when $b > 0$). Distributions stay roughly constant. It is only for extremely high productivity shocks that the firm finds it optimal to tap costly external equity financing. In this figure we also plot the right-hand side of the collateral constraint (3.20). As explained in DeAngelo, DeAngelo, and Whited (2011), the firm prefers to stay below its debt capacity. The intuition is that costly external equity implies that the firm optimally preserves debt capacity so that it can have sufficient slack to take advantage of future productivity shocks.

Finally, we conduct comparative statics exercises, such as those in DeAngelo, DeAngelo, and Whited (2011). For these exercises, which can be found in Figure 18, we use the parameterization used to construct Figure 14, except that we set the risk free rate to 0.01 so that we end up with realistic leverage ratios. The panels in this figure depict average leverage and average debt capacity utilization as functions of the various model parameters. Debt capacity utilization is described as the ratio of debt to the collateral constraint given by (3.20). We plot debt capacity utilization in order to highlight the effect of the Lagrange multiplier in (3.24) on optimal leverage policy. As emphasized in DeAngelo, DeAngelo, and Whited (2011), this effect can be quite strong.

The main piece of intuition behind this figure is that lower optimal debt accompanies any model parameter that increases the probability of needing external equity finance (equivalently, of getting too close to debt capacity) or that increases the cost of equity finance. This intuition is highlighted in the first two panels, which depict leverage policy as a function of the parameters that govern the cost of external finance. The striking result is that a higher cost of equity finance *lowers* optimal leverage and decreases debt capacity utilization. This model prediction is exactly opposite of what one would find in a typical static model, in which firms substitute away from high cost and toward low cost financial instruments. The reason for the seemingly counterintuitive result in the dynamic model is that these static models cannot embody the option value of maintaining financial flexibility. In other words, in this sort of dynamic models, as equity financing becomes more costly, the firm foregoes the tax benefits of debt in favor of “keeping its powder dry” by lowering leverage in order to take advantage of possible future investment opportunities.

In general, this last effect is striking. In the baseline simulation, the average ratio of debt to debt capacity is only 0.63, because optimal policy only dictates exhausting debt capacity when the firm encounters a series of high productivity shocks. In the simulation the firm rapidly pays down the debt after it has funded its investment projects. In general, we only see high debt capacity utilization when equity finance is “free” or when optimal investment policy rarely requires outside funding. For example, firms with low volatility or low serial correlation shocks or firms with smooth investment tend to take advantage of the tax benefit of debt and have high leverage (full debt capacity utilization), whereas firms with lumpier or more volatile investment conserve debt capacity and keep leverage optimally low.

We conclude this section by describing further results in the literature that have been produces by this class of models. Hennessy and Whited (2005) contains two further important results. First, the paper provides explanations for the stylized fact that leverage is negatively correlated with lagged measures of profitability.[24] This fact appears to stand in the face of static trade-off theory, which predicts that highly profitable firms should lever up to shield their profits from taxes. However, in a dynamic trade-off model with endogenous investment, highly profitable firms are likely to have low leverage because they are more likely to have financing surpluses. They therefore simply do not need leverage to fund optimal investment. In addition, they are likely to resemble the type of firm in Figure 15, whose marginal benefit of debt is low.

Second, Hennessy and Whited (2005) offers an explanation for the stylized fact, documented in Baker and Wurgler (2002) that leverage varies negatively with a variable called lagged weighted q. This variable is the weighted average of the most recent values of a firm’s Tobin’s q, with more weight given to those observations accompanied by security issuance. Baker and Wurgler

[24]See, for example, Rajan and Zingales (1995).

(2002) interpret this finding as indicative of firms' attempts to time the equity issuance market. In contrast, the model in Hennessy and Whited (2005) shows that firms with high productivity shocks simultaneously have high qs and finance high desired investment with equity. Increased investment raises next period's firm value. The debt to assets ratio then falls and stays low because of leverage serial correlation. Therefore, in simulated data, one observes a negative correlation between current low leverage and those values of Tobin's q that were accompanied by security issuance.

3.5 Cash and Debt

One of the drawbacks of the model in Section 3.4 is that it can only really make predictions about *net* debt because of the assumption that cash is negative debt. Therefore, the next layer of complication that one can add to this class of models is a firm that can simultaneously hold both cash, p, and debt, b. Gamba and Triantis (2008) study this model. The central insight is that without any transactions costs associated with issuing debt, the firm never optimally holds both debt and cash. However, as long as the firm incurs costs when it *changes* its level of debt, then it can be optimal for the firm to have positive stocks of debt and cash simultaneously. The intuition behind this insight is the same intuition behind the model from Section 3.3: firms hold cash to avoid having to resort to costly external finance. The main difference here is that external finance can take the form either of debt or of equity.

To outline this model, we let $q(b', b)$ represent the costs associated with debt issuance. In this case, the sources and uses of funds identity can be written as:

$$\begin{aligned} e(k, k', p, p', b, b', z) \equiv (1-\tau)\pi(k, z) - \psi(k' - (1-\delta)k, k) - (k' - (1-\delta)k) \\ + \frac{b'}{1 + r(1-\tau)} - b - q(b', b) - \frac{p'}{1 + r(1-\tau)} + p. \end{aligned} \tag{3.25}$$

Once again, the effects of adding both cash and debt are seen primarily in the sources and uses of funds identity (3.25). The rest of the dynamic maximization problem is largely similar, except that the firm chooses three variables, cash, debt, and capital, instead of just two.

As in the case of the model in Hennessy and Whited (2005), the model in Gamba and Triantis

(2008) is substantially more complex. It too has a personal taxation and fire sales of assets when the firm has insufficient internally generated funds to repay debt. The existence of these fire sales increases the value of cash by making external finance in the form of debt potentially more costly. The main result in Gamba and Triantis (2008) is that when the firm experiences low profitability, if it wants to decrease its net debt position, it should increase its cash balance rather than paying down debt. The reason is that the cash investment is costlessly reversible, whereas paying down debt is not. Conversely, if the firm wished to increase its net debt position, it is usually optimal for it to do so via manipulation of its cash levels. These results in turn imply that, ceteris paribus, two firms with the same net debt level can have very different valuations, with the one with more cash, and therefore more flexibility, usually being more valuable.

3.6 Risky Debt

One of the less attractive features of the model we have built thus far is that debt is risk-free. Several papers in the literature have relaxed this assumption. In the finance literature Moyen (2004), Hennessy and Whited (2007), and Titman and Tsyplakov (2007) have used what are called "endogenous default" models to study and quantify, respectively, the effects of financial frictions on investment cash flow sensitivity, the cost of external finance, and the effects of agency conflicts on leverage dynamics.[25]

To outline this type of model, we revert back to the assumption that cash equals negative debt, drop the state variable, p, and allow the state variable, b, to assume both positive and negative values. We also drop the collateral constraint (3.20) so that the firm can take on as much debt as it desires. The sources and uses of funds identity then looks very similar to (3.21), with two small changes:

$$\begin{aligned} e(k,k',b,b',z) \equiv\ & (1-\tau)\pi(k,z) - \psi(k'-(1-\delta)k,k) - (k'-(1-\delta)k) + \frac{b'}{1+\tilde{r}(z,k',b')} \\ & + \frac{\tau\tilde{r}(z,k',b')b'}{(1+\tilde{r}(z,k',b'))(1+r)} - b \end{aligned} \tag{3.26}$$

[25]Related models in the macroeconomics literature include Pratap and Rendon (2003) and Cooley and Quadrini (2001), which examine the effects of financial frictions on capital accumulation and industry dynamics, respectively.

The first change is that the interest rate on debt is a risky rate, $\tilde{r}(z, k', b')$, which a function of the current shock and the firm's next-period choice of capital and debt. To see why the risky interest rate depends on these three variables, first note that in these kinds of models the firm optimally defaults when its equity value equals zero. Thus, the lender knows the default states corresponding to each possible (k', b', z') triple. The firm chooses (k', b'), so that given the current shock, z, the lender also knows the probability distribution over the default states corresponding to the firm's (k', b') choice. The risky interest rate is then set so that banks earn zero expected profits, and it is therefore a function of (z, k', b').

The second change is the term $(\tau\tilde{r}(z, k', b')b')/(1+\tilde{r})/(1+r)$, which implies that the firm takes the present value of the interest tax deduction in the period in which it issues debt. Although clearly not what happens in reality, this feature greatly simplifies the determination of the risky interest rate. Otherwise, the tax deduction would depend on *last* period's shock, and current profits would then depend on *four* state variables, which would include last period's shock.

If the firm defaults, the lender recovers the firm's profits and assets less any deadweight default costs, α, so that total recovery is $R(k, z) \equiv (1-\alpha)((1-\tau)\pi(k, z) + (1-\delta)k)).$[26] At this point the firm is left to start over again. The endogenous risky interest rate is then set so that the expected rate of return on debt equals the risk free rate, given that the bank receives less than 100% recovery in default. This zero-profit condition is

$$b'(1+r) = \int_{\text{default states}} R(k', z')dg(z', z) + b' \int_{\text{solvent states}} (1+\tilde{r})dg(z', z) \tag{3.27}$$

The Bellman equation for the problem can then be written as:

$$V(k, b, z) = \max\left\{0, \max_{k', b'}\left\{e(k, k', b, b', z) + \eta(e(k, k', b, b', z)) + \frac{1}{1+r}\int V(k', b', z')dg(z' \mid z)\right\}\right\}, \tag{3.28}$$

This expression captures the default option by expressing equity value as the maximum of zero

[26]In Hennessy and Whited (2007) the lender also extracts all of the bargaining surplus from the firm, which equals the (opposite of) current period cash flow that would produce what equity value would be if the firm did not have limited liability.

and the equity value of the firm as a going concern. Although the Bellman equation appears almost identical to (3.22), the firm's problem is more complicated because it interacts with a competitive lender to set the risky interest rate. To solve the model, one needs to know firm value in order to calculate the default states and thus the interest rate. However one needs to know the interest rate to calculate firm value. One feasible solution algorithm is a "loop-within-a-loop," which proceeds as follows. First, assume that the interest rate on debt is the risk-free rate and solve for the value function. With this estimate of the value function, one can calculate the default states and the interest rate, which is a function of the model's state variables. Then, using this new interest rate, one repeats this procedure until the value function converges.[27] The complexity of these models is one disadvantage they have with respect to contingent claims models, in which default thresholds can often be found analytically and can almost always be characterized as a differential equation.

These models convey much of the intuition from the simpler risk-free debt model of Section 3.4. In particular, optimal leverage declines with the serial correlation and variance of the productivity shock, as well as with profit function curvature and the cost of external equity finance. Notwithstanding these similarities, this type of model contains richer intuition than collateral-constraint models. In particular, leverage is not constrained by available collateral. Instead, firms optimally keep leverage at a moderate level to avoid default and the consequent deadweight default costs. Of course, as the deadweight costs increase, leverage falls. Less obviously, because default occurs when equity value is zero, optimal leverage increases with any technological feature that helps the firm more efficiently transform capital into equity value. For example, a high capital depreciation rate lowers average leverage because a high depreciation rate implies that any given amount of capital creates less value. This last piece of intuition cannot be gleaned from simpler models in which the size of the capital stock is the main determinant of debt capacity. Thus, the direct effects of value on debt capacity is unique to endogenous default models

Moyen (2004) uses this type of a model to reexamine the question from Gomes (2001) and Alti (2003) of whether the sensitivity of investment to cash flow is a good measure of finance

[27]See Gomes and Schmid (2010) and Moyen (2004) for other algorithms.

constraints. To this end, the paper compares two models: a full-blown endogenous default model with costly external finance and an endogenous risky interest rate and an analogous model in which all sources of outside financing have been shut down. Moyen (2004) finds that shutting off external finance lowers firm value and reduces investment cash flow sensitivity. The intuition is that firms that cannot access external finance have limited investment opportunities. As in the simplest investment model, even conditioning on Tobin's q, cash flow is a proxy for investment opportunities because of production function concavity. When faced with constraints on external finance, the firm invests less aggressively when hit with a positive cash flow shock.

This effect operates in Hennessy and Whited (2007) as well, with regard to increasing the cost of external equity financing. The intuition with regard to increasing the deadweight costs of bankruptcy is different. In the face of high deadweight default costs, the firm hoards financial assets, thereby allowing it to respond strongly to shocks to cash flow. Thus investment cash flow sensitivity rises with the deadweight costs of default.

Hennessy and Whited (2007) also attempt to estimate the costs of external finance by performing a structural estimation. The goal is to infer these costs by examining firm behavior through the lens of the model. The main result is that there exist large indirect costs of external equity that exceed simple underwriting fees. In addition, deadweight bankruptcy cost estimates are in line with previous studies.

3.7 Other Models

The goal of this section has been to provide a strong basis for understanding the workings of dynamic models of financing and investment. In so doing, we have not provided an exhaustive survey of all of the papers that have used these kinds of models. This section attempts to clean up these loose ends.

Dynamic models of investment and financing are a natural place to start understanding the related areas of corporate diversification and restructuring. For example, the model in Gomes and Livdan (2004) features a firm that can expand into an additional industry at a fixed cost. The main result is that firms optimally expand when their original division becomes unproductive, so that it is worth it to the firm to pay the fixed cost of diversifying. Thus, the widely documented "diversification discount" emerges from the model as a natural consequence of diminishing returns to scale. Warusawitharana (2008) examines the related question of asset purchases and sales, which, in contrast to acquisitions, generate substantial shareholder value. Using a related model, he shows that these gains are an endogenous outcome of firm value maximization.

The area of behavioral finance related to market timing has recently seen studies that use dynamic investment models. For example, Bolton, Chen, and Wang (2012) essentially add stochastic costs of equity issuance to the model in Bolton, Chen, and Wang (2011) to understand when issuing overvalued stock or repurchasing undervalued stock is optimal. One of many interesting results is that firms use market timing to alleviate financial constraints and can thus smooth investment much more than they would in the absence of any market timing ability. Another paper in this area is Warusawitharna and Whited (2012), which models actual equity misvaluation instead of stochastic issuance costs. This paper is also a structural estimation paper instead of a pure theory paper. Their main conclusions are that managers do appear to time the market and add value for long-term shareholders. However, this issuance and repurchasing activity has more of an effect on firm financial policies than on real investment policies.

One area of corporate finance that has seen little formal dynamic research is payout policy. We are aware of one recent paper that fits loosely into the class of models treated in this section: Lambrecht and Myers (2012). Risk-averse, habit-forming managers maximize their utility, which is a function of rents. Rents, in turn, are a residual after the firm invests, borrows, and distributes funds to shareholders. There are no taxes or financial frictions in the model, so that Modigliani-Miller holds. However, managers are subject to a threat of intervention by outside shareholders. Managers therefore pay out just enough dividends to keep their jobs. This threat ties dividends to managerial rents. In this setting the original partial adjustment model from Lintner (1956) arises from relatively primitive assumptions about preferences. The main result is that risk-aversion causes managers to smooth rents over time. Because dividends are tied to rents via the control threat, managers also smooth dividends. Managers' habit formation also causes rents and payouts only partially to adjust to changes in permanent income.

Finally, other literatures besides corporate finance have used this class of models to understand a wide variety of phenomena. The first is asset pricing. At this point several papers have employed dynamic models of financing and investment to try to understand phenomena ranging from the effects of finance constraints on expected returns (Whited and Wu 2006; Li, Livdan, and Zhang 2009; Livdan, Sapriza, and Zhang 2009), the low expected returns of highly levered firms (Gomes and Schmid 2010), and credit risk (Kuehn and Schmid 2011). The second literature is macroeconomics. In the wake of the 2007–2009 financial crisis, studies such as Covas and Den Haan (2011) and Jermann and Quadrini (2012) have used versions of these models to understand the cyclical properties of leverage and how financial shocks affect real activity.

3.8 Summary

We close this section by pointing out that although dynamic models of investment and financing are rich in their treatment of dynamic effects, they are less rich in their treatment of fundamental reasons behind the existence of financial frictions. For example, in these model debt and equity issuance costs are specified exogenously, so that the firm is powerless to influence its own cost of external finance. In addition, the form of financial contracts (usually equity and one-period debt) is exogenous. Finally, most (but not all) of these models are implicitly specified under the risk-neutral measure, which implies that one cannot disentangle the effects of risk from the effects of financial frictions on corporate policies.

Of course, many of these simplifications have been made because the purpose of these models is not to explain why financial frictions exist but to understand the consequences of financial frictions. In addition, many of these simplifications have been adopted because one of the main purposes of these models is to serve as a vehicle for using actual data to quantify economic phenomena. There is a sharp trade-off between the feasibility of taking a model directly to the data and the stringency of the assumptions regarding financial contracts. A few studies such as Hennessy, Livdan, and

Miranda (2010) and Schmid (2011) have made some progress in trying to understand whether these simplifications have important quantitative effects; however, further work would be interesting.

4 Structural Estimation

One feature of all of the studies thus far reviewed that we have yet to discuss is whether and how the models can be estimated. Some of these studies, such as Gomes (2001), Moyen (2004), and Gamba and Triantis (2008) are pure theoretical pieces. Others, such as Hennessy (2004) and Riddick and Whited (2009) contain both models and reduced form empirical evidence. Still others, such as Hennessy and Whited (2005, 2007) have a great deal of empirical content in the form of structural estimation exercises. We therefore turn to the next broad topic—structural estimation.

Structural estimation is an attempt to fit a model directly to data, to assess the quality of the fit, to identify parameters that govern technology, preferences, and (thus far in corporate finance) largely time-invariant institutional features. In particular, structural estimation ascertains whether optimization models generate data that resemble data from real-world firms. As such, structural estimation is an exercise in using a realistic theoretical structure to interpret the data. Estimating models is *useful* because it allows estimation of parameters that can be used to quantify the primitives that shape firm behavior. Thus, parameter estimates allow us to measure quantities that, as financial economists, we find interesting, such as the cost of external finance or managerial preferences over shirking. These parameters can also be used to study counterfactual situations that can be useful for policy evaluation. Estimating models is also useful because the connection between theory and tests of theory is extremely tight, thereby allowing a transparent interpretation of any results. In contrast, interpretation of many reduced-form regressions is more difficult because the assumptions of the underlying, often verbal, model are not spelled out.

Interestingly, structural estimation may or may not require a dynamic—as opposed to a static—model. However, most modern incarnations of structural estimation, at least since Hansen and Singleton (1982), have employed dynamic optimization models to generate equations to be estimated. In corporate finance structural estimation starts with Whited (1992), but despite this almost 20 year history, the total number of structural estimation papers in corporate finance is tiny. Prominent examples include Bond and Meghir (1994), Love (2003), Hennessy and Whited (2005, 2007), Sorensen (2007), DeAngelo, DeAngelo, and Whited (2011), Kang, Liu, and Qi (2010), Lin, Ma, and Xuan (2011), Taylor (2010, 2012), Nikolov and Whited (2012), Albuquerque and Schroth (2010), Morellec, Nikolov, and Schürhoff (2012), Matvos and Seru (2011), Gantchev (2011), Glover (2011), Page (2012), Korteweg (2010), Dimopoulos and Sacchetto (2011) and Coles, Lemmon, and Meschke (2012), Schroth, Suarez, and Taylor (2012).

As in the case of theoretical dynamic models, we conjecture that the paucity of studies stems from the perception that structural estimation is too complex, so that lack of transparency makes it difficult to learn anything from these sorts of exercises. Therefore, one goal of this section is to explain the intuition behind these techniques to dispel this perception of complexity. Once again, the intent is not to provide mathematical rigor, which can be found in any econometrics textbook, but to provide insight into how these techniques work and "advice from the trenches" designed to help researchers avoid common pitfalls.

The second goal of this section is therefore not only to review the literature, which we structure around the four main methods that have been used so far in corporate finance for structural estimation: generalized method of moments (GMM), simulated method of moments (SMM), and simulated maximum likelihood, (SMLE). We then survey the literature. We close with a comparison between calibration and structural estimation, and with a brief overview of the wealth of other structural estimation techniques that have not yet been widely applied in corporate finance.

4.1 GMM and Euler Equations

The papers using GMM, starting in corporate finance with Whited (1992), typically estimate investment Euler equations, which are variants of (3.8).[28] In reviewing this literature, we do not review the basics of GMM, which can be found in any graduate econometrics textbook. Instead, we look at assumptions necessary to apply GMM to specific problems.

[28]See Bond and Meghir (1994), Love (2003), Kang, Liu, and Qi (2010), Lin, Ma, and Xuan (2011), and Liu, Whited, and Zhang (2009). One important example that does not involve investment Euler equations is Albuquerque and Schroth (2010), which uses exactly identified GMM on a static model of a takeover premium.

Hansen and Singleton (1982) point out that estimating *any* Euler equation for any dynamic decision requires an assumption of rational expectations. This assumption allows the empirical researcher to replace the unobservable expected cost of delaying investment in estimating equations such as (3.8) with the observable realized cost of delaying investment plus an expectational error. The intuition behind this replacement is straightforward: as a general rule, what happens is equal to what one expects plus one's mistake. Further, the mistake has to be orthogonal to any information available at the time that the expectation was made; otherwise, the expectation would have been different. This last observation implies that lagged endogenous variables are orthogonal to the expectational error and that they therefore can be used as instruments to estimate Euler equations.[29]

One useful feature of GMM is that it is accompanied by a general specification test—the test of the over identifying restrictions. Of course, the validity of this test relies on the assumption that one has as many instruments as parameters, and this assumption (also known as an exclusion restriction) cannot be tested. Finally, it is worth reemphasizing that one should have "strong" instruments. In the case of nonlinear GMM, instrument strength cannot be cast in terms of the degree of correlation between the instruments and endogenous regressors. Instead, it is usually cast as full rank of the gradient of the moment conditions with respect to the parameters. Wright (2003) offers a useful diagnostic test of instrument strength for nonlinear models. Euler estimation with panel data can be accomplished with a variety of different statistical packages, such as Stata or SAS, so that implementing this kind of estimation is straightforward.

We now turn to a brief review of this literature. Whited (1992) estimates a nonlinear investment Euler equation with data on U.S. firms, finding that the overidentifying restrictions of the model are not rejected for financially healthy firms but that they are rejected for financially unhealthy firms. By parameterizing a Lagrange multiplier on a constraint that limits firm financing (for example, such as (3.20)), and by substituting this parameterization for the Lagrange multiplier in the Euler

[29]It is worth pointing out that using lagged instruments on a generic regression does not constitute structural estimation, which specifically requires that an economic model directly produce the estimating equation. In the case of Euler equations, the model also produces the assumptions allowing the validity of lagged instruments.

equation, the paper also finds that financially constrained firms tend to behave as if they have a higher discount rate than their financially unconstrained counterparts; that is, they postpone investment until the future. This result uncovers a specific economic effect of financing constraints, and it could not have been found in a static regression because the concept of delay is by definition a dynamic concept.

Kang, Liu, and Qi (2010) and Lin, Ma, and Xuan (2011) conduct similar exercises in order to understand, respectively, the effect of the Sarbanes-Oxley Act on corporate investment and the effect of ownership structure on financial constraints and thus investment. In contrast, Bond and Meghir (1994) and Love (2003) linearize the Euler equation. They demonstrate using models such as that in Section 3.4 that financial variables enter into the specification in the presence of financial frictions such as an increasing interest rate schedule. They then straightforwardly test for the significance of these variables. They conclude that financial frictions are important for investment.

4.2 Simulated Method of Moments

To motivate this section, we note that one serious drawback with estimating the first-order conditions from dynamic models is the assumptions one needs to make in order to produce a closed-form estimating equation. For example, investment Euler equations make the assumption of quadratic investment adjustment costs, which, as explained in Section 3.1.3, imply that the firm optimally wants to smooth investment over time. This assumption is at odds with the lumpiness often observed in the time-series of firm investment. More realistic models that relax this unpalatable assumption, however, require much more computationally intensive techniques if one wants to estimate them.

To sum up, there exists a tension between realism and the sorts of models that can produce closed-form estimating equations. Better models that can explain more phenomena may not lend themselves to closed-form solutions or to smooth, differentiable first-order conditions. Fortunately, the development of the econometrics of simulation estimators, such as Pakes and Pollard (1989), Ingram and Lee (1991), Duffie and Singleton (1993), and Gourieroux, Monfort, and Renault (1993), has allowed researchers to bring better models to the data. Although many different simulation estimators are used in other fields such as industrial organization, macroeconomics, and labor economics, we review the two main types that have been used in corporate finance. The first is simulated method of moments (SMM) and the second is simulated maximum likelihood. The rest of this section outlines how to do an SMM estimation and reviews the literature that uses this technique.

4.2.1 Outline

SMM is conceptually simple. As detailed in Section 3.1.2, discrete-time dynamic models typically have a solution in the form of a rule that prescribes optimal policy tomorrow, given the firm's state today (such as its levels of debt and capital) and a random shock received today. This rule can be used to generate a panel of simulated data by simulating a series of shocks and tracing out the firm's optimal choices. The researcher then calculates interesting moments using both the simulated data and a real data set and then methodically searches for model parameters that make the simulated moments as close as possible to their corresponding real moments. Estimation requires at least as many moments as model parameters. Examples of parameters in the models in Section 3 include profit function curvature, the serial correlation and variance of the z shock, etc.

An SMM estimation usually proceeds in two separate set of steps.

First Set of Steps:

1. Choose a set of moments to match. The list can include means, variances, covariances, regression coefficients, and so on, but one must have at least as many moments as parameters to estimate.

2. Calculate these moments with real data and stack them in a vector. We denote the estimated vector of real moments as $M(x)$, in which x is an *i.i.d.* data sample.

3. Calculate the covariance matrix of the moments. Invert it. This matrix is the GMM weight matrix. We denote the estimated weight matrix as W.

Second Set of Steps:

1. Pick starting values for the set of parameters, which we call β_0.

2. Using β_0, solve the model, create simulated data using the policy function, and calculate the same moments that were calculated with the real data. Stack them in a vector. We denote the simulated moments as $m(y, \beta)$, in which y is a simulated data sample. Note that the simulated moments are a function of the parameter vector, β, and can change when β changes.

3. Form the SMM moment vector as the stack of simulated moments minus the stack of real moments.

4. Form the SMM objective function exactly as one would form a GMM objective function, that is, as a weighted sum of squared errors, in which the "errors" are given by the moment vector, and the weights are given by the weight matrix. Let the objective function be denoted as $Q(x, y, \beta)$:

$$Q(x, y, \beta) \equiv (M(x) - m(y, \beta))' \, W \, (M(x) - m(y, \beta)) \,. \tag{4.1}$$

5. Find the parameter vector, $\hat{\beta}$, that minimizes the SMM objective function.

6. Adjust the standard errors and test statistics for simulation error. Let J be the ratio of the number of observations in the simulated data set to the number observations in the real data set, which we denote as N. Then the covariance matrix of $\sqrt{N}(\hat{\beta} - \beta)$ is given by:

$$\left(1 + \frac{1}{J}\right) \left((\partial m(y, \beta)/\partial \beta)' \, W \, (\partial m(y, \beta)/\partial \beta)\right)^{-1} \tag{4.2}$$

The term $1 + 1/J$ is the adjustment for simulation error, which approaches 1 as $J \longrightarrow \infty$. The rest is a standard GMM covariance matrix.

7. Use either a general test of the overidentifying restrictions of the model as a specification test or, as in DeAngelo, DeAngelo, and Whited (2011), test the equality of the moments individually. The test of the overidentifying restrictions can be written as:

$$\frac{NJ}{1+J}Q(x, y, \beta) \tag{4.3}$$

The t-statistics on the individual moment conditions are calculated exactly as pricing errors in a standard GMM framework. See, for example, Cochrane (2005).

From this recipe, it is clear that the "data" part of SMM can be done separately from any computations involving the model. Thus, the hardest part of doing SMM is solving the economic model. Once one is armed with a solution, it is straightforward to simulate data and calculate simulated moments. Nonetheless, SMM can be time intensive because one has to solve the model with every step of the econometric minimization routine.

4.2.2 Identification

The success of SMM relies fundamentally on picking moments that can identify the structural parameter vector, β. "Identification" means something different in this context than it does in an instrumental variables setting, so it is useful to elaborate on this difference. In an IV setting, "identification" typically means finding a source of exogenous variation in the data. This exogenous variation allows one to ascribe a causal interpretation to any regression coefficients one obtains. In other words, because IV or OLS can fit lines through most data scatters, identification in this context can be thought of as making enough assumptions about the source of data variation to interpret estimated elasticities as causal relations. In other words, it amounts to finding a suitable lens through which to interpret the data. In structural estimation, finding a suitable lens involves writing down a model, and the word "identification" is used in the econometric sense that an econometric objective function have a unique minimum.

Both the mathematical and intuitive conditions for identification of an SMM estimator are similar to those for GMM. We focus on the intuitive aspect, which requires that the moments be informative about the structural parameters, or equivalently, that the sensitivity of the moments to the parameters be high. This requirement is obviously analogous to the importance of using high quality instruments in a standard IV framework, that is, instruments highly correlated with the endogenous regressors.

Ideally, each parameter would be identified by a single moment. Changing that parameter would cause that moment, and only that moment, to vary. Of course, this hurdle is very high and not actually necessary. In fact, if it were, then an economic model probably would not be necessary. What is much more likely to occur can be intuitively described as staggered identification, which is what we exploit whenever we solve linear systems of equations. A simple example helps illustrate how identification works. Suppose we are using three moments (a, b, c) to identify three parameters (x, y, z), and that the economic model implies the following moment conditions:

$$\begin{aligned} a &= x + y \\ b &= y + z \\ c &= x + z \end{aligned}$$

If we know a, b and c, we can uniquely identify the (x, y, z), even though each moment is affected by several parameters and each parameter affects several moments. SMM exploits this same feature in a non-linear way.

How does one ensure that the model is identified? First, check the parameter standard errors. If one has chosen bad moments, the standard errors will be large because the sensitivity of the moments to the parameters enters into the calculation of the standard errors. Second, and similarly, if the estimation does not converge, the model parameters are likely not identified.

Third, and most important, the best way to pick good moments is to understand the economic forces that drive the model, just as the best way to pick good instruments is to understand the economics of the question being asked. The economic intuition behind SMM thus comes from the identification conditions, which can only come from a careful understanding of the model. Understanding the economics of a model that has no closed form solution requires time because it requires conducting comparative statics exercises and plotting moments versus parameters. The ideal moment has a steep, monotonic relation to a parameter, and moments should move in different directions for different parameters. Ensuring meaningful identification thus entails an attempt to disentangle which parameters affect which moments. Finally, it is important not to "cherry pick" moments. A model should be able to make sense of many features of the data, and it is useful and often instructive to determine where the model does and does not do a good job of matching the data.

To understand the importance of identification, we construct a simple counterexample of an *unidentified* parameter as follows. Suppose one were to add operating leverage to the simple model from section 3.1 and remove capital adjustment costs so that the sources and uses of funds identity (3.1) becomes

$$e(k, k', z) \equiv \pi(k, z) - (k' - (1 - \delta)k) - \varphi k, \tag{4.4}$$

in which the term, φk, is operating leverage. It is straightforward to see that it impossible to identify the parameters δ and φ separately using data on the capital stock. It is only possible to estimate their sum, $\delta + \varphi$. However, if one were to find separate data on capital depreciation, it might be possible to identify φ using data on profitability.

To close our discussion of identification, we add one word of caution. Just as there does not exist any perfectly exogenous source of data variation in observational studies, structural estimation does not magically solve all endogeneity problems. Structural estimation accounts for any endogeneity within the model. However, just as any linear econometric model suffers from omitted variables problems to one degree or another, there will always be elements omitted from an estimated structural model. Therefore, one needs to worry about whether omitting those elements biases the parameter estimates. If one understands the model one is estimating, this type of important thought experiment is usually feasible, and it often opens doors for future research.

4.2.3 Practical Advice

Several pieces of practical advice are in order. The first is to use real data to estimate the weight matrix. In both GMM and SMM, the weight matrix is the inverse of the covariance matrix of the moment conditions. In this type of an SMM exercise, the weight matrix is therefore just the covariance matrix of the data moments alone. This weight matrix (roughly) puts the most weight on the most precisely estimated moments.

It is not necessary to use any model simulations to estimate this covariance matrix. Instead, it can be estimated by calculating all of the moments jointly as a large GMM system. Another, often computationally simpler, method is to calculate the moments separately, calculate the influence function for each moment, stack the influence functions of the moments, and take their inner product.[30] Heuristically speaking, an influence function of an estimator is a function of the data whose mean has the same asymptotic distribution as the estimator. For example, suppose an *i.i.d.* sample of a random variable, x, has a sample average $\bar{x}$. Then the influence function is trivially $x-\bar{x}$. See Newey and McFadden (1994) for more precise definitions of influence functions. One final important issue is accounting for time-series dependence in the data when estimating the weight matrix. In panel data, this issue is usually accomplished via a clustering algorithm.

The second piece of advice is not to use an identity matrix as a weight matrix because it mechanically puts the most weight on the moment that is largest in absolute value. This advice lies in contrast to that given in Cochrane (2005) regarding tests of asset pricing models. The difference is that returns on different portfolios are all of the same order of magnitude. In contrast, moments used in a corporate finance simulated moments exercise can be of very different magnitudes.

The third piece of advice is a set of small pointers concerning the minimization of the SMM objective function. It is computationally infeasible to use gradient based methods such as Newton-Raphson. One can use a grid search if one is only estimating one or two parameters. For more complex problems, other alternatives include various linear-programming algorithms such as Nelder-Meade or simulation based algorithms such as Metropolis-Hastings or the closely related simulated annealing. As a general rule, simulation based algorithms do a good job of efficiently getting close to a global minimum, but their final convergence is often glacial. On the other hand, linear-programming based algorithms work much more quickly but are also more likely to get stuck in a local minimum. Therefore, a useful approach is to start with a simulation based algorithm and

[30]See Erickson and Whited (2012).

then switch to a linear-programming algorithm when the estimation is near convergence. A formal switching rule can be cast in terms of a loose convergence criterion. Next, it is important to pick a good starting value for the minimization algorithm. This choice is often a byproduct of comparative statics exercises that one should do. It is equally important to try several starting values to increase the probability of finding a global minimum of the SMM objective function. Next, one must use the same pseudo-random draws for each simulation of the artificial data. Otherwise, it is possible to attribute a change in the SMM objective function to a change in a parameter, when it is actually due to a change in the pseudo-random draws. Finally, simulate data sets that are several times as large as the size of one's actual data set. The simulation step is computationally cheap, so it makes sense to lower one's standard errors as much as possible by using large simulations to remove as much simulation error as possible.

The fourth piece of advice concerns heterogeneity. The models we have considered here are usually of a single firm or at best of an industry equilibrium, in which heterogeneity comes from realizations of shocks. Thus, SMM estimates the parameters of an average firm, not the average parameter across firms. These two quantities are equal in linear models, but not necessarily in the nonlinear models considered here. However, corporate finance data are generated by heterogeneous firms in different industries with different technologies, different sources of uncertainty, and differential access to financial markets. Thus, interpretation of any results is easiest on relatively homogeneous subpopulations in firms. For example, DeAngelo, DeAngelo, and Whited (2011) estimate their model on different industries. Whether this type of sample splitting is feasible or not, it is nonetheless important to extract as much heterogeneity from the data as possible, usually by removing fixed effects. Often in estimating dynamic models, it is natural to use autoregressive coefficients as moments. In this case, one cannot, of course, demean data on a firm by firm basis. A feasible alternative is the double differencing method in Han and Phillips (2010).

4.2.4 Model Design

The most important piece of advice concerns model design. It is essential to write down a realistic model that can shed light on the questions one wants to answer. To illustrate this point, we turn to the careful model design in Taylor (2010). This paper asks three related questions. It asks why so few CEOS are fired every year (2%). It also asks how much firing one should expect from a well-functioning board, and how much shareholder value is destroyed if boards do not function well. The second two questions are particularly well suited to structural estimation because they require calculating a counterfactual, that is, asking what would happen if the model parameters were different from the estimated parameters.

The paper then postulates four possible reasons for the low firing rate. The costs to shareholders of turning over the CEO may be large. If the next best CEO is as good as the current one, why bother? Boards can learn slowly about CEO ability. CEOs can be entrenched. What makes this paper useful is that the model is tailored specifically to answer these questions.

The model can be summarized as follows. In each period the board decides whether or not to fire the CEO. The CEO also retires or quits with a certain probability, and the firm generates profits. These profits are the sum of three components: industry profits, firm-specific profits, and CEO turnover costs if the CEO is fired. Firm specific profitability can be affected by the CEO's ability, but the board cannot distinguish the industry and firm-specific components of profits.

New CEOs are drawn from a talent pool characterized by normally distributed ability. The board of directors has priors over this distribution, so that when the CEO starts, the board attributes mean CEO talent to the CEO. However, the board updates its beliefs about CEO ability each period when it observes profitability. The board then maximizes its utility, which consists of the expected discounted value of future profit flows minus any personal costs incurred if the board fires the CEO. The board optimally fires the CEO as soon as posterior mean skill drops below an endogenous threshold

The model's parameters capture the four possible reasons for firing CEOs. This feature of the design of the model implies that estimating these parameters sheds light on the importance of the different reasons. The model contains an entrenchment parameter that takes the form of the board's personal costs (loss of a golf partner) of firing the CEO. It contains a parameter quantifying the costs to shareholders of CEO turnover. Lack of better possible replacement CEOs is measured by the variance of the CEO talent pool. Finally, slow learning on the part of the board is captured by the statistical properties of the profit signals that the board does observe.

The empirical results are interesting. The parameter estimates indicate that boards have high personal costs of firing CEOs—so high that boards behave as if firing a CEO costs shareholders 5.9% of assets, whereas it really only costs them 1.3% of assets. Also, in counterfactual experiments that set the board's personal costs to zero, the model predicts that 13% of CEOs would be fired every year if boards were to act in shareholders' interests.[31] This type of counterfactual exercise is one of the main reasons for doing structural estimation because counterfactuals that revolve around preferential differences are difficult to do in any other way.

4.3 Simulated Maximum Likelihood

Very recently, a new methodology has been introduced into structural estimation in empirical corporate finance: simulated maximum likelihood. Simulated maximum likelihood comes in many different flavors, so we focus on the particular flavor that has recently been used in corporate finance in Morellec, Nikolov, and Schürhoff (2012) (MNS, hereafter).

To begin explaining the technique used here in general terms, suppose one has written down an *economic* model that can produce the density of a variable, x, of interest. This density is completely characterized by a parameter β and can thus be denoted as $f(x_i \mid \beta)$, in which x_i is an observation from the random sample, x. One example of such a density comes from a real options model, such as the one outlined in Section 2.2, in which the marginal product of capital fluctuates between two bounds. The firm invests in a lumpy fashion when the marginal product hits a bound, thus driving the marginal product back to a return point. Another example can be found in the dynamic capital structure model in Section 2.4, in which cash flows (and hence leverage) fluctuate between two bounds. For many models in this class, the density is usually piecewise exponential.

Of course, if one has a closed-form solution for the density $f(x \mid \beta)$, it is possible to estimate the model parameters via standard MLE. However, because one has a closed-form density, it is possible

[31] Of course, some entrenchment may be optimal for shareholders ex ante if, for example, it improves the pool of potential new CEOs.

to add parameter heterogeneity to the *econometric* model. The ability to address heterogeneity directly is one important advantage of simulated MLE over moment-based estimators. One adds heterogeneity by expressing β as $\beta_i \equiv \beta + u_i$, in which u_i is a random variable that varies across the observations in the data set. In this case the density becomes $f(x \mid \beta, u_i)$.

If one is willing to make distributional assumptions about u_i, such as $u_i \sim f(u_i)$, then one can integrate u_i out of the density to arrive at the marginal density of x. In this kind of a mixture model, it is rare that one can derive a closed form solution for the marginal density. Instead, one can use Monte Carlo integration as follows. Simulate H random draws, u_i^h, from $f(u_i)$ and approximate the integral by:

$$\frac{1}{H}\sum_{j=1}^{h} f\left(x_i, \beta \mid u_i^h\right) f\left(u_i^h\right) \tag{4.5}$$

Now the SMLE estimator is defined as:

$$\hat{\beta} \equiv \arg\max_{\beta} \sum_{i=1}^{N} \ln\left(\frac{1}{H}\sum_{h=1}^{H} f\left(x_i, \beta \mid u_i\right) f\left(u_i^h\right)\right), \tag{4.6}$$

and the usual MLE distribution theory applies. Unlike an SMM estimator, which is consistent even for small H, the SMLE is consistent only if *both* N and H go to infinity. Thus, it is important to use many simulated observations in the Monte Carlo integration.

Thinking about identification is just as important here as it is for SMM. The technical conditions for identification of an SMLE estimator are identical to those for an MLE estimator. Heuristically, one cannot have a flat likelihood around the true parameter value. This requirement implies that the likelihood has to change shape when a parameter moves, and it has to change shape in different ways for different parameters. Therefore, just as in the case of SMM, it is important to do many comparative statics exercises to determine whether one's likelihood does in fact change shape with different parameter values.

MNS use this technique to estimate a contingent claims model of leverage that is closely related to the model in Morellec (2004), in which managers can steal cash flows. Shareholders can only stop managers by conducting costly control challenges; so managers steal just enough to keep the shareholders from ousting them. Without agency problems, as explained in Section 2, this class of models often produces leverage ratios that are too high relative to those observed in the real world. The tax-advantages of debt loom large in comparison to the expected bankruptcy costs. The converse of this problem is that excessively high debt issuance costs are required to match model generated leverage with real world leverage. The reason is that equity values drift upward, thereby mechanically causing leverage to drift downward. With high issuance costs, firms restructure their debt infrequently, and the downward drift keeps model and real leverage in line. In MNS, however, the distribution of optimal model-generated leverage ratios (including the mean) can be reconciled with the distribution of actual leverage ratios primarily because managers keep leverage low in order to be able to divert cash flows to themselves rather than to bondholders.

The estimation starts with a derivation of the distribution of leverage in the model. At this point MNS add parameter heterogeneity as described above. First, they assume that each estimated model parameter has a normal distribution. Therefore, for each possible set of model parameters, leverage must have a different distribution. This joint distribution is intractable, but they solve this problem by doing maximum likelihood with the marginal distribution of leverage, which they calculate by using Monte Carlo integration to integrate out the parameter heterogeneity. This last step is what adds a numerical dimension to their estimation problem.

MNS get two benefits from all of this added complexity. First, they quantify managers' average perceived costs of control (private benefit from resource diversion) to be approximately 2% of firm value. These kinds of quantitative predictions are hard to make in the absence of a theoretical lens through which to view the data. Second, the addition of random effects, u_i allows them to see whether parameter estimates vary in sensible ways with observed firm characteristics, which is akin to an out-of-sample test. They find that the most important determinant of managerial leeway to steal is institutional ownership.

This kind of exercise is much less time consuming than SMM, and the problem of picking moments to match disappears if one is maximizing a likelihood. However, one does need to be able to derive a closed form likelihood function from one's model. This requirement puts restrictions on the classes of models that can be estimated with this technique. Further, these types of models tend to be harder to identify because one needs to extract a great deal of information from the distribution of one variable, in this case, leverage. Thus SMM and SMLE have different advantages and disadvantages, and the choice of technique is likely to be dictated by the model being estimated.

4.4 Literature Review

In the process of describing the different techniques that can be used for structural estimation, we have reviewed several of the extant studies. However, we have also omitted many more, and the intent of this section is to summarize the rest of the literature. First, as detailed in Section 4.1, the literature on investment Euler equations has been almost exclusively been interested in the effects of some sort of financial friction on investment. The subsequent literature that uses more computationally intensive techniques has been able, in contrast, to address a much wider range of interesting questions. Thus, although Hennessy and Whited (2005, 2007) and DeAngelo, DeAngelo, and Whited (2011) examine the kinds of capital structure and financial frictions questions on which we have focused, other studies have ventured into closely related areas such as cash holdings, as well as very different areas such as shareholder activism and executive compensation.

In the broad area of capital structure, Nikolov and Whited (2012) estimate a dynamic model of investment and cash via SMM to understand the extent to which agency issue affect corporate cash holdings. Their main result from this estimation is that agency issues related to self dealing are more important for explaining corporate cash balances but that agency issues related to firm size are more important for firm value.

Again in capital structure, Korteweg (2010) uses Markov Chain Monte Carlo methods to estimate the net benefits to debt, that is, the tax benefits less financial distress costs. The general challenge in this literature is to isolate financial distress costs from economic distress that usually leads to financial distress. The idea inKorteweg (2010) to derive estimating equations from robust identities for levered and unlevered betas and a linear leverage model, which can be derived from the Leland (1994) model. These equations express the benefits of leverage as a function of the level of leverage, and the the benefit beta as a function of debt and equity betas. These equations, along with the assumption that similar firms have identical asset betas allow the recovery of the unobservable net benefits. Bayesian estimation techniques are useful for this exercise because they easily allow estimation in the presence of missing data. The main results are that the maximum net benefit for the median firm is about 5% of value, and that the average firm is under levered. However, the second result mostly comes from the zero leverage firms in the sample.

One further structural estimation study dealing with financial distress is Glover (2011). This paper is based on the clever observation that using observed defaults to estimate the costs of financial distress, understates these costs substantially. The reason is that firms with high costs optimally choose low leverage and thus do not default. The only way to tackle this self selection problem is to have an estimate of optimal leverage, and one feasible way to estimate this quantity is with a model. Glover (2011) uses SMM to estimate a variant of the contingent claims model in Goldstein, Ju, and Leland (2001). His main result is that firms expect to lose 45% of their value in default—a number much higher than previous estimates.

Executive compensation has also been home to structural estimation studies. As detailed above Taylor (2010) uses structural estimation of a dynamic learning model to understand why CEOs are seldom fired. Taylor (2012) uses estimation of a related model to ask a different question: how does CEO pay respond to news about CEO ability. One of the more interesting results from this exercise is that CEOs have downwardly rigid pay when bad news arrives. Using a model helps make sense of this result. It does not appear to be driven by weak governance. Instead, the model estimates imply that CEOs accept low pay in return for the insurance provided by downward wage rigidity. Page (2012) asks the converse question of why firm value appears to be insensitive to changes in CEO *incentive* (as opposed to total) pay.[32] To answer the question, he estimates a model in which a risk averse CEO and a board interact to set the CEO's contractual incentive pay. The main message from the estimation is that performance is insensitive to pay because CEO's have an extremely high marginal utility of leisure; that is, they are already working very hard.

[32]See also Coles, Lemmon, and Meschke (2012).

Structural estimation has also been used to understand takeovers. For example, Albuquerque and Schroth (2010) estimate the private benefits of control by examining the discounts and premia associated with block trades. The intuition starts with a model in which that block trades transfer private benefits of control and that the block trades at a premium when the owner can fend off a possible subsequent tender offer. Therefore, Albuquerque and Schroth (2010) can extract estimates of the private benefits of control from the premia and discounts. The main result is that private benefits represent about 3% of equity value. Dimopoulos and Sacchetto (2011) tackles a related question by trying to understand whether large takeover premia represent takeover resistance or initial preemptive bidding on the part of one of several potential acquirers. To separate these two possible explanations, Dimopoulos and Sacchetto (2011) estimate an auction model of takeovers that encompasses both explanations. The most important result from this estimation is that target resistance is the main determinant of takeover premia.

Several papers in a smorgasbord of other areas of corporate finance have also used structural estimation. Sorensen (2007) asks whether better venture capitalists (VCs) are more likely to take their entrepreneurs public. This question is difficult because good entrepreneurs end up getting matched with better VCs. This self selection problem cannot be treated with a standard Heckman correction because the matching between VCs and entrepreneurs is not a zero-one decision. Instead Sorensen (2007) uses Markov Chain Monte Carlo methods to estimate a two-sided matching model of VCs and entrepreneurs. The results from this estimation can then be used to correct for the sample selection bias in the estimation of the effect of VC quality on entrepreneurial outcomes. The main result is that a standard profit estimation of the effect of VCs on IPO probabilities is overstated by a factor of two. Gantchev (2011) estimates a sequential decision model of a hedge fund's decision to conduct a proxy contest. The intent is to understand whether the returns earned by shareholder activists cover their monitoring costs. Interestingly, the mean net activist return is near zero, but the top quartile of activists earn substantial net returns. Schroth, Suarez, and Taylor (2012) estimate a dynamic model of bank runs to measure the fragility that results from financing long-term assets using dispersed, short-term debt. The estimates of the model then allow evaluation of various policies designed to prevent bank runs. The main finding is that interventions targeting asset liquidity and conduit leverage are most effective. Warusawitharana (2011) estimates a dynamic model in which innovation improves firm productivity in order to understand variation in R&D intensity across industries.

4.5 Calibration versus Estimation

We now comment on the differences between structural estimation and an exercise called "calibration," which is similar to SMM inasmuch as calibration also tries to match model-generated stylized facts with stylized facts in the data. Both exercises are useful, but they serve different purposes, so it is important to keep them separate. From a purely technical point of view, calibration is different from structural estimation mostly because model calibrations are not accompanied by standard errors for the model parameters, and structural estimations do provide standard errors. The reason is that calibration only tries to match a few stylized facts with many model parameters, and whenever there are too many degrees of freedom, inference is impossible. In contrast, SMM matches at least as many stylized facts as model parameters, and therefore standard errors can be calculated in the same way as GMM standard errors. Calibration is especially useful in situations in which model estimation is infeasible but in which one nonetheless wants to learn from the model.

However, the distinction between calibration and estimation goes deeper. The main purpose of calibration is to ensure that a model with a numerical solution provides (usually directional) predictions that are likely to be empirically relevant. This type of exercise can be extremely helpful for rounding out intuition. For example, Strebulaev (2007) picks parameters for his model in a variety of ways: primarily by matching moments and distributions from different data sets, and by using estimates from previous studies. As discussed above, the calibration in this paper advances understanding of extant empirical results by running regressions on simulated data. Thus, although calibration is not an empirical exercise, it can yield useful results as a part of a theoretical exercise.

In contrast, structural estimation tries to find the unique set of model parameters that reconcile the model with the data. It is therefore an explicitly *empirical* exercise that attempts to "stress test" a single model to compare the performance of competing models. Any SMM estimation that uses more moments than parameters is an example of stress testing. If one uses enough moments, *any* model will fail, but it is useful to know where. As an example of model comparisons, Whited (1992) compares investment models with and without financing frictions. Structural estimation also is useful as an empirical exercise because it can be used to measure quantities that are of interest to financial economists, such as the cost of external financing (Hennessy and Whited 2007).

4.6 Other Methods

The three methods here are those that have been used in empirical corporate finance. A wealth of other methods exists in asset pricing, macroeconomics, labor economics, and industrial organization. For example, macroeconomics and asset pricing researchers use a method closely related to SMM called indirect inference (Gourieroux, Monfort, and Renault 1993). The idea behind this technique is that it is often impossible to characterize the likelihood of the data from a dynamic model. However, it is often possible to use what is called an "auxiliary" model to approximate the likelihood. As long as the mapping between the model parameters and the parameters of the auxiliary model is unique, one can then estimate the structural parameters of a model by minimizing the difference between the parameters of the auxiliary model estimated on real data and the parameters of the auxiliary model estimated on model simulated data.

The empirical industrial organization literature contains examples of many different techniques. For example, one can estimate the parameters of a dynamic discrete choice model by constructing a likelihood function directly from the model, as in the seminal work of Rust (1987). Other methods sidestep the necessity of solving the model in the first place. For example, Hotz and Miller (1993) show that one can estimate the parameters of a dynamic discrete choice model using a two-step method. First, one estimates, nonparametrically, the choice probabilities from the data and then one uses these estimates to back out the model parameters from the model without having to solve it repeatedly. Aguirregabiria and Mira (2007) build upon this idea to formulate a method for estimating dynamic discrete games that have multiple equilibria. One final, but by no means exhaustive, example is in Bajari, Benkard, and Levin (2007), which uses an estimated policy function in conjunction with the optimality conditions of the model to estimate the structural model parameters.

At this point, we are aware of only two examples of the use of these methods in corporate finance. Matvos and Seru (2011) uses the methods in Bajari, Benkard, and Levin (2007) to examine the extent to which inefficient resource allocation contributes to the diversification discount. Kang, Lowery, and Wardlaw (2012) uses the methods in Hotz and Miller (1993) to estimate the direct and indirect costs to taxpayers of bailing out failed banks. Clearly, the scope is enormous for incorporating these methods into corporate finance to answer interesting questions.

5 Conclusion

This survey has attempted to cover the broad area of dynamic corporate finance. One lesson that emerges is that the field is extraordinarily diverse. For example, we have covered three broad areas—contingent claims models, discrete-time financing and investment models, and structural estimation—and each area is markedly different in terms of the types of analytical devices it employs. Structural estimation can be performed on either class of theoretical models, but the two classes of theoretical models are distinct. Thus, at this point, a brief comparison is in order.

Contingent claims models offer two important advantages over discrete-time investment models. First, they allow the pricing of claims, being based on techniques developed in the derivatives *pricing* literature. Investment-based models, in contrast, have only just started to be used to price claims via the addition of a pricing kernel (Gomes and Schmid 2010). Second, dynamic contingent claims models allow for unbounded firm growth because of the homogeneity property that allows large firms to be thought of as scaled up versions of small firms. Investment-based models, in contrast, study firm behavior in a bounded set of possible values for the state variables. Thus, while they are extremely useful for studying ratios (leverage ratios, investment rates, etc.), they are less useful for studying long-term growth, as they can, at best, be interpreted as representing fluctuations around long-term deterministic trends.

Discrete-time financing and investment models also have a unique set of advantages. In these models, investment is endogenous and the firm faces a period-by-period sources and uses of funds identity. This set-up allows the study of several phenomena that are outside the realm of contingent claims models. For example, it is possible to investigate the joint dynamics of investment and finance—a concept that is not well-defined in a contingent claims model. Further, it is possible to examine the interplay between a pecking-order-like incentive to use the cheapest source of funds versus the classic trade-off between tax benefits and distress costs. Both of these motives operate in a dynamic investment and financing model, thus giving rise to behavior that resembles neither a pecking order nor a static trade-off. Only the trade-off motive operates in a contingent claims model. Finally, it is possible to model explicitly an array of financing vehicles that is larger than just simple "debt" and "equity," such as distributions to shareholders and cash.

These two classes of models constitute different types of theory. In contrast, structural estimation is a specific type of empirical exercise that fits theoretical (as opposed to statistical) models directly to data. In corporate finance, structural estimation has mainly been used on the types of models surveyed here, but the only requirement for estimation is that the model parameters can be identified with the relevant data. The main advantage of structural estimation is that interpreting any empirical results—even the sign of a regression coefficient—usually requires a model. Structural estimation puts the model first and makes the model explicit, and thus the results are often simpler to interpret than the results from reduced form regressions, which are often the product of a verbal model.

Further work would be interesting in other areas as well. For largely historical reasons, the areas of investment, leverage, and finance constraints have seen the bulk of research in dynamic corporate finance. As noted above, very little work has been done on payout, and the studies that have tackled issues in executive compensation are few (e.g., Taylor 2010, 2012; Brisley 2006; He 2009; Noe and Rebello 2012). Other areas that have seen scant work include product markets (Lambrecht 2001, 2004; Morellec and Zhdanov 2008) and entrepreneurship (Wang, Wang, and Yang 2012). In general, any area of corporate finance could be amenable to dynamic models and structural estimation.

Our essay should be seen in the proper light. In our review of dynamic corporate finance we do not argue against conventional modeling apparatus, or worse, intuition developed in the process of working on such models. We are very much in favor of non-technical economic intuition and consider it to be of utmost importance. On the one hand, simple static models are useful for understanding economic mechanisms in their simplest forms and getting benchmark results. On the other hand, every dynamic model, however complicated it may seem, should be explained by straightforward economic intuition. Ability to explain intuition to one's grandmother is a non-trivial test of scientific inquiry in social sciences. Moreover, we certainly do not claim that there is no benefit from simple intuitive thinking without any formal models. Instead, we believe it should be, and most often is, the starting stage for more complicated and realistic models.

We also do not argue against conventional reduced form or quasi-experimental empirical methods. Instead, we wish to explain how structural methods can add to our understanding of corporate finance questions in new and interesting ways. In particular, parameter estimates obtained from structural estimations are useful for counterfactual (i.e. "what-if") analysis that can be used at a minimum to further our understanding of how firms respond to primitives or at best to evaluate policies. In addition, the number of high-quality natural experiments available to corporate finance researchers is likely much smaller than the number of interesting questions to be asked, so that structural estimation often offers a more feasible alternative for understanding these interesting questions.

Our biggest claim is twofold. First, however penetrating static models can sometimes be, they are only a starting block in our understanding of any phenomenon. The first question that should be asked is whether the results are robust in a dynamic world. Second, pure qualitative judgment cannot take us too far in our desire to distinguish between various economic mechanisms, and therefore a closer link between modeling framework and empirical methods should be established. In a way, we should not only take our models more seriously, and expect more from them, but also be able to criticize them with better precision.

References

Abel, Andrew B., and Janice C. Eberly, 1994, A unified model of investment under uncertainty, *American Economic Review* 84, 1369–1384.

Acharya, Viral, Sergei A. Davydenko, and Ilya A. Strebulaev, 2012, Cash holdings and credit risk, *Review of Financial Studies* forthcoming.

Aguirregabiria, Victor, and Pedro Mira, 2007, Sequential estimation of dynamic discrete games, *Econometrica* 75, 1–53.

Albuquerque, Rui, and Enrique Schroth, 2010, Quantifying private benefits of control from a structural model of block trades, *Journal of Financial Economics* 96, 33–55.

Alti, Aydoḡan, 2003, How sensitive is investment to cash flow when financing is frictionless?, *Journal of Finance* 58, 707–722.

Altinkilic, Oya, and Robert S. Hansen, 2000, Are there economies of scale in underwriter spreads? Evidence of rising external financing costs, *Review of Financial Studies* 13, 191–218.

Anderson, Ronald W., and Andrew Carverhill, 2012, Corporate liquidity and capital structure, *Review of Financial Studies* 25, 797–837.

Anderson, Ronald W., and Suresh Sundaresan, 1996, Design and valuation of debt contracts, *Review of Financial Studies* 9, 37–68.

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

Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin, 2007, Estimating dynamic models of imperfect competition, *Econometrica* 75, 1331–1370.

Baker, Malcolm, and Jeffrey Wurgler, 2002, Market timing and capital structure, *Journal of Finance* 57, 1–32.

Berk, Jonanthan, and Peter DeMarzo, 2011, *Corporate Finance* (Pearson).

Bhamra, Harjoat S., Lars-Alexander Kuehn, and Ilya A. Strebulaev, 2010a, The aggregate dynamics of capital structure and macroeconomic risk, *Review of Financial Studies* 23, 4187–4241.

———, 2010b, The levered equity risk premium and credit spreads: A unified framework, *Review of Financial Studies* 23, 645–703.

Black, Fisher, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 1981, 637–654.

Bolton, Patrick, Hui Chen, and Neng Wang, 2011, A unified theory of tobin's q, corporate investment, financing, and risk management, *Journal of Finance* 66, 1545–1578.

———, 2012, Market timing, investment, and risk management, *Journal of Financial Economics* forthcoming.

Bond, Stephen, and Costas Meghir, 1994, Dynamic investment models and the firm's financial policy, *Review of Economic Studies* 61, 197.

Boyle, Glenn W., and Graeme A. Guthrie, 2003, Investment, uncertainty, and liquidity, *Journal of Finance* 58, 2143–2166.

Bradley, Michael, Gregg A. Jarrell, and E. Han Kim, 1984, On the existence of an optimal capital structure: Theory and evidence, *Journal of Finance* 39, 857–878.

Brander, James A., and Tracy R. Lewis, 1986, Oligopoly and financial structure: The limited liability effect, *American Economic Review* pp. 956–970.

Brennan, Michael J., and Eduardo S. Schwartz, 1985, A new approach to evaluating natural resource investments, *Midland Corporate Finance Journal* 3, 37–47.

Brisley, Neil, 2006, Executive stock options: Early exercise provisions and risk-taking incentives, *Journal of Finance* 61, 2487–2509.

Broadie, Mark, Mikhail Chernov, and Suresh Sundaresan, 2007, Optimal debt and equity values in the presence of chapter 7 and chapter 11, *Journal of Finance* 62, 1341–1377.

Campello, Murillo, and Dirk Hackbarth, 2012, The firm-level credit multiplierq, *Journal of Financial Intermediation* 21, 446–472.

Carlson, Murray, and Ali Lazrak, 2010, Leverage choice and credit spreads when managers risk shift, *Journal of Finance* 65, 2323–2362.

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

———, Jianjun Miao, and Neng Wang, 2010, Entrepreneurial finance and nondiversifiable risk, *Review of Financial Studies* 23, 4348–4388.

Childs, Paul D., David C. Mauer, and Steven H. Ott, 2005, Interactions of corporate financing and investment decisions: The effects of agency conflicts, *Journal of Financial Economics* 76, 667–690.

Childs, Paul D., Steven H. Ott, and Alexander J. Triantis, 1998, Capital budgeting for interrelated projects: A real options approach, *Journal of Financial and Quantitative Analysis* 33.

Cochrane, John H., 2005, *Asset Pricing: (Revised)* (Princeton University Press: Princeton, NJ).

Coles, Jeffrey L., Michael L. Lemmon, and J. Felix Meschke, 2012, Structural models and endogeneity in corporate finance: The link between managerial ownership and corporate performance, *Journal of Financial Economics* 103, 149 – 168.

Cooley, Thomas F., and Vincenzo Quadrini, 2001, Financial markets and firm dynamics, *American Economic Review* 91, 1286–1310.

Cooper, Russell, and Joao Ejarque, 2003, Financial frictions and investment: Requiem in q, *Review of Economic Dynamics* 6, 710–728.

Covas, Francis, and Wouter J. Den Haan, 2007, Cyclical behavior of debt and equity using a panel of canadian firms, Manuscript, Bank of Canada.

Covas, Francisco, and Wouter J. Den Haan, 2011, The cyclical behavior of debt and equity finance, *American Economic Review* 101, 877–899.

Danis, András, Daniel Rettl, and Toni M. Whited, 2012, Testing dynamic tradeoff theory: Evidence from rebalancing points, Manuscript University of Rochester.

Davydenko, Sergei, Ilya A. Strebulaev, and Xiaofei Zhao, 2012, A market-based study of the cost of default, *Review of Financial Studies* 25, 2959–2999.

Davydenko, Sergei A., 2010, What triggers default? a study of the default boundary, Manuscript, University of Toronto.

DeAngelo, Harry, Linda DeAngelo, and Toni M. Whited, 2011, Capital structure dynamics and transitory debt, *Journal of Financial Economics* 99, 235–261.

DeAngelo, Harry, and Ronald W. Masulis, 1980, Optimal capital structure under corporate and personal taxation, *Journal of Financial Economics* 8, 3–29.

Décamps, Jean-Paul, Thomas Mariotti, and Stéphane Villeneuve, 2005, Investment timing under incomplete information, *Mathematics of Operations Research* 30, 472–500.

———, 2009, Investment timing under incomplete information: Erratum, *Mathematics of Operations Research* 34, 255–256.

Dimopoulos, Theodosios, and Stefano Sacchetto, 2011, Preemptive Bidding, Target Resistance, and Takeover Premiums, Manuscript, Carnegie Mellon University.

Dixit, Avinash, 1989, Entry and exit decisions under uncertainty, *Journal of Political Economy* pp. 620–638.

Dixit, Avinash K., and Robert S. Pindyck, 1994, *Investment under Uncertainty* (Princeton University Press: Princeton, NJ).

Duffie, Darrell, and Kenneth J Singleton, 1993, Simulated moments estimation of markov models of asset prices, *Econometrica* 61, 929–52.

Dumas, Bernard, 1991, Super contact and related optimality conditions, *Journal of Economic Dynamics and Control* 15, 675–685.

Erickson, Timothy, and Toni M. Whited, 2012, Treating measurement error in Tobin's q, *Review of Financial Studies* 25, 1286–1329.

Fama, Eugene F., and Kenneth R. French, 2002, Testing trade-off and pecking order predictions about dividends and debt, *Review of Financial Studies* 15, 1–33.

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

Fazzari, Steven M., R. Glenn Hubbard, and Bruce C. Petersen, 1988, Financing constraints and corporate investment, *Brookings Papers on Economic Activity* 1988, 141–206.

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

Franks, Julian R., and Walter N. Torous, 1989, An empirical investigation of us firms in reorganization, *Journal of Finance* pp. 747–769.

Gamba, Andrea, and Alexander Triantis, 2008, The value of financial flexibility, *Journal of Finance* 63, 2263–2296.

Gantchev, Nickolay, 2011, The costs of shareholder activism: Evidence from a sequential decision model, Manuscript, University of North Carolina.

Glover, Brent, 2011, The expected cost of default, Manuscript, Carnegie Mellon University.

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

Gomes, Joao, and Dmitry Livdan, 2004, Optimal diversification: Reconciling theory and evidence, *Journal of Finance* 59, 507–535.

Gomes, Joao F., 2001, Financing investment, *American Economic Review* 91, 1263–1285.

——— , and Lukas Schmid, 2010, Levered returns, *Journal of Finance* 65, 467–494.

Gorbenko, Alexander S., and Ilya A. Strebulaev, 2010, Temporary versus permanent shocks: Explaining corporate financial policies, *Review of Financial Studies* 23, 2591–2647.

Gourieroux, Christian S., Alain Monfort, and Eric Renault, 1993, Indirect inference, *Journal of Applied Econometrics* 8, S85–S118.

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

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

——— , and Campbell R. Harvey, 2001, The theory and practice of corporate finance: Evidence from the field, *Journal of Financial Economics* 60, 187–243.

Grenadier, Stecen R., 1995, Valuing lease contracts a real-options approach, *Journal of Financial Economics* 38, 297–331.

——— , 2002, Option exercise games: An application to the equilibrium investment strategies of firms, *Review of Financial Studies* 15, 691–721.

Grenadier, Steven R., and Andrey Malenko, 2010, A Bayesian approach to real options: The case of distinguishing between temporary and permanent shocks, *Journal of Finance* 65, 1949–1986.

Grenadier, Steven R., and Neng Wang, 2005, Investment timing, agency, and information, *Journal of Financial Economics* 75, 493–533.

——— , 2007, Investment under uncertainty and time-inconsistent preferences, *Journal of Financial Economics* 84, 2–39.

Grenadier, Steven R., and Allen M. Weiss, 1997, Investment in technological innovations: An option pricing approach, *Journal of Financial Economics* 44, 397–416.

Guo, Xin, Jianjun Miao, and Erwan Morellec, 2005, Irreversible investment with regime shifts, *Journal of Economic Theory* 122, 37–59.

Hackbarth, Dirk, 2009, Determinants of corporate borrowing: A behavioral perspective, *Journal of Corporate Finance* 15, 389–411.

———, 2010, Managerial traits and capital structure decisions, *Journal of Financial and Quantitative Analysis* 43, 843.

———, Christopher A. Hennessy, and Hayne E. Leland, 2007, Can the trade-off theory explain debt structure?, *Review of Financial Studies* 20, 1389–1428.

Hackbarth, Dirk, Richmond Mathews, and David Robinson, 2012, Capital structure, product market dynamics, and the boundaries of the firm, Manuscript, Duke University.

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

Hackbarth, Dirk, and Jianjun Miao, 2012, The dynamics of mergers and acquisitions in oligopolistic industries, *Journal of Economic Dynamics and Control* 36, 585–609.

———, and Erwan Morellec, 2006, Capital structure, credit risk, and macroeconomic conditions, *Journal of Financial Economics* 82, 519–550.

Hackbarth, Dirk, and Erwan Morellec, 2008, Stock returns in mergers and acquisitions, *Journal of Finance* 63, 1213–1252.

Han, Chirok, and Peter C. B. Phillips, 2010, GMM estimation for dynamic panels with fixed effects and strong instruments at unity, *Econometric Theory* 26, 119–151.

Hansen, Lars Peter, and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* pp. 1269–1286.

Hayashi, Fumio, 1982, Tobin's marginal q and average q: A neoclassical interpretation, *Econometrica* 50, 213–224.

He, Zhiguo, 2009, Optimal executive compensation when firm size follows geometric brownian motion, *Review of Financial Studies* 22, 859–892.

Hennessy, Christopher A., 2004, Tobin's q, debt overhang, and investment, *Journal of Finance* 59, 1717–1742.

———, Amnon Levy, and Toni M. Whited, 2007, Testing Q theory with financing frictions, *Journal of Financial Economics* 83, 691–717.

Hennessy, Christopher A., Dmitry Livdan, and Bruno Miranda, 2010, Repeated signaling and firm dynamics, *Review of Financial Studies* 23, 1981–2023.

Hennessy, Christopher A., and Toni M. Whited, 2005, Debt dynamics, *Journal of Finance* 60, 1129–1165.

———, 2007, How costly is external financing? Evidence from a structural estimation, *Journal of Finance* 62, 1705–1745.

Hotz, V. Joseph, and Robert A. Miller, 1993, Conditional choice probabilities and the estimation of dynamic models, *Review of Economic Studies* 60, 497.

Ingram, Beth F., and Bong-Soo Lee, 1991, Simulation and estimation of time series models, *Journal of Econometrics* 47, 197–205.

Jermann, Urban, and Vincenzo Quadrini, 2012, Macroeconomic effects of financial shocks, *American Economic Review*, forthcoming.

Kane, Alex, Alan J. Marcus, and Robert L. McDonald, 1984, How big is the tax advantage to debt?, *Journal of Finance* 39, 841–853.

Kang, Ari, Richard Lowery, and Malcolm Wardlaw, 2012, Bank failures and regulatory incentives: A structural estimation, Manuscript, University of Texas, Austin.

Kang, Qiang, Qiao Liu, and Rong Qi, 2010, The Sarbanes-Oxley act and corporate investment: A structural assessment, *Journal of Financial Economics* 96, 291–305.

Kim, E. Han, 1989, Optimal capital structure in Miller's equilibrium, in Sudipto Bhattacharya, and George Constantinides, ed.: *Financial Markets and Incomplete Information* . pp. 36–48 (Rowman and Littlefield: Totowa, NJ).

Klein, Manuel, 2009, Comment on "Investment timing under incomplete information", *Mathematics of Operations Research* 34, 249–254.

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, *Journal of Finance* 65, 2137–2170.

——— , and Ilya A. Strebulaev, 2012, An (s, s) model of capital structure: Theory and evidence, Manuscript, Stanford University.

Kuehn, Lars Alexander, and Lukas Schmid, 2011, Investment-based corporate bond pricing, Manuscript, Duke University.

Kurshev, Alexander, and Ilya A. Strebulaev, 2006, Firm size and capital structure, Manuscript, Stanford University.

Lambrecht, Bart M., 2001, The impact of debt financing on entry and exit in a duopoly, *Review of Financial Studies* 14, 765–804.

——— , 2004, The timing and terms of mergers motivated by economies of scale, *Journal of Financial Economics* 72, 41–62.

——— , and Stewart C. Myers, 2007, A theory of takeovers and disinvestment, *Journal of Finance* 62, 809–845.

——— , 2012, A Lintner model of payout and managerial rents, *Journal of Finance* 67, 1761–1810.

Leland, Hayne E., 1994, Corporate debt value, bond covenants, and optimal capital structure, *Journal of Finance* 49, 1213–1252.

Li, Erica X. N., Dmitry Livdan, and Lu Zhang, 2009, Anomalies, *Review of Financial Studies* 22, 4301–4334.

Lin, Chen, Yue Ma, and Yuhai Xuan, 2011, Ownership structure and financial constraints: Evidence from a structural estimation, *Journal of Financial Economics* 102, 416–431.

Lintner, John, 1956, Distribution of incomes of corporations among dividends, retained earnings, and taxes, *American Economic Review* 46, 97–113.

Liu, Laura Xiaolei, Toni M. Whited, and Lu Zhang, 2009, Investment-based expected stock returns, *Journal of Political Economy* 117, 1105–1139.

Livdan, Dmitry, Horacio Sapriza, and Lu Zhang, 2009, Financially constrained stock returns, *Journal of Finance* 64, 1827–1862.

Love, Inessa, 2003, Financial development and financing constraints: International evidence from the structural investment model, *Review of Financial Studies* 16, 765–791.

Lucas, Robert E., 1978, Asset prices in an exchange economy, *Econometrica* 46, 1429–1446.

———, and Edward C Prescott, 1971, Investment under uncertainty, *Econometrica* 39, 659–81.

Mankiw, N. Gregory, 1985, Small menu costs and large business cycles: A macroeconomic model of monopoly, *Quarterly Journal of Economics* 100, 529–537.

Matvos, Gregor, and Amit Seru, 2011, Resource allocation within firms and financial market dislocation: Evidence from diversified conglomerates, manuscript.

McDonald, Robert, and Daniel Siegel, 1986, The value of waiting to invest, *Quarterly Journal of Economics* 101, 707.

McDonald, Robert L., and Daniel R. Siegel, 1985, Investment and the valuation of firms when there is an option to shut down, *International Economic Review* 26, 331–349.

Mehra, Rajnish, and Edward C. Prescott, 1985, The equity premium: A puzzle, *Journal of Monetary Economics* 15, 145–161.

Mella-Barral, Pierre, and William Perraudin, 1997, Strategic debt service, *Journal of Finance* 52, 531–556.

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

Miao, Jianjun, 2005, Optimal capital structure and industry dynamics, *Journal of Finance* 60, 2621–2659.

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

Miranda, Mario J., and Paul L. Fackler, 2002, *Applied Computational Economics and Finance* (MIT Press: Cambridge, MA).

Modigliani, Franco, and Merton H. Miller, 1958, The cost of capital, corporation finance and the theory of investment, *American Economic Review* 48, 261–297.

———, 1963, Corporate income taxes and the cost of capital: A correction, *American Economic Review* 53, 433–443.

Morellec, Erwan, 2001, Asset liquidity, capital structure, and secured debt, *Journal of Financial Economics* 61, 173–206.

———, 2004, Can managerial discretion explain observed leverage ratios?, *Review of Financial Studies* 17, 257–294.

———, Boris Nikolov, and Norman Schürhoff, 2012, Corporate governance and capital structure dynamics, *Journal of Finance* 67, 803–848.

Morellec, Erwan, and Norman Schürhoff, 2010, Dynamic investment and financing under personal taxation, *Review of Financial Studies* 23, 101–146.

Morellec, Erwan, and Norman Schürhoff, 2011, Corporate investment and financing under asymmetric information, *Journal of Financial Economics* 99, 262–288.

Morellec, Erwan, and Alexei Zhdanov, 2008, Financing and takeovers, *Journal of Financial Economics* 87, 556–581.

Moyen, Nathalie, 2004, Investment-cash flow sensitivities: Constrained versus unconstrained firms, *Journal of Finance* 59, 2061–2092.

Myers, Stewart C., and Nicholas S. Majluf, 1984, Corporate financing and investment decisions when firms have information that investors do not have, *Journal of Financial Economics* 13, 187–221.

Newey, Whitney, and Daniel McFadden, 1994, Large sample estimation and hypothesis testing, in Robert Engle, and Daniel McFadden, ed.: *Handbook of Econometrics, Vol. 4* . pp. 2111–2245 (North-Holland: Amsterdam).

Nikolov, Boris, and Toni M. Whited, 2012, Agency conflicts and cash: Estimates from a structural model, Manuscript University of Rochester.

Noe, Thomas H., and Michael J. Rebello, 2012, Optimal corporate governance and compensation in a dynamic world, *Review of Financial Studies* 25, 480–521.

Page, Beau, 2012, CEO ownership and firm value: Evidence from a structural estimation, Manuscript, University of Rochester.

Pakes, Ariel, and David Pollard, 1989, Simulation and the asymptotics of optimization estimators, *Econometrica* 57, 1027–1057.

Pawlina, Grzegorz, 2010, Underinvestment, capital structure, and strategic debt restructuring, *Journal of Corporate Finance* 16, 679–702.

Pratap, Sangeet, and Silvio Rendon, 2003, Firm investment in imperfect capital markets: A structural estimation, *Review of Economic Dynamics* 6, 513–545.

Rajan, Raghuram G., and Luigi Zingales, 1995, What do we know about capital structure? Some evidence from international data, *Journal of Finance* 50.

Riddick, Leigh A., and Toni M. Whited, 2009, The corporate propensity to save, *Journal of Finance* 64, 1729–1766.

Rust, John, 1987, Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher, *Econometrica* pp. 999–1033.

Sarkar, Sudipto, 2000, On the investment-uncertainty relationship in a real options model, *Journal of Economic Dynamics and Control* 24, 219–225.

———, 2003, The effect of mean reversion on investment under uncertainty, *Journal of Economic Dynamics and Control* 28, 377–396.

Schaefer, Stephen M., and Ilya A. Strebulaev, 2008, Structural models of credit risk are useful: Evidence from hedge ratios on corporate bonds, *Journal of Financial Economics* 90, 1–19.

Schmid, Lukas, 2011, A quantitative dynamic agency model of financing constraints, Manuscript, Duke University.

Scholes, Myron S., and Mark A. Wolfson, 1992, *Taxes and Business Strategy: A Planning Approach* (Prentice Hall: Englewood Cliffs, NJ).

Schroth, Enrique, Gustavo Suarez, and Lucian A. Taylor, 2012, Dynamic debt runs: Evidence from a structural estimation, Manuscript, University of Pennsylvania.

Sorensen, Morten, 2007, How smart is smart money: A two-sided matching model of venture capital, *Journal of Finance* 62, 2725–2762.

Stokey, Nancy L., 2009, *The Economics of Inaction: Stochastic Control Models with Fixed Costs* (Princeton University Press: Princeton, NJ).

———, Robert E. Lucas, and Edward C. Prescott, 1989, *Recursive Methods in Economic Dynamics* (Harvard University Press: Cambridge, MA).

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

———, Haoxiang Zhu, and Pavel Zryomov, 2012, Dynamic information asymmetry, financing, and investment, Manuscript, Stanford University.

Summers, Lawrence, H., 1987, Investment incentives and the discounting of depreciation allowances, in Martin Feldstein, ed.: *The Effects of Taxation on Capital Accumulation* (Chicago University Press: Chicago).

Tauchen, George, 1986, Finite state markov-chain approximations to univariate and vector autoregressions, *Economics Letters* 20, 177–181.

———, and Robert Hussey, 1991, Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models, *Econometrica* 59, 371–396.

Taylor, Lucian A., 2010, Why are CEOs rarely fired? evidence from structural estimation, *Journal of Finance* 65, 2051–2087.

——— , 2012, CEO pay and CEO power: Evidence from a dynamic learning model, *Journal of Financial Economics* forthcoming.

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

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

Tsyplakov, Sergey, 2008, Investment frictions and leverage dynamics, *Journal of Financial Economics* 89, 423–443.

Wang, Chong, Neng Wang, and Jinqiang Yang, 2012, Dynamics of entrepreneurship under incomplete markets, Journal of Financial Economics, forthcoming.

Warusawitharana, Missaka, 2008, Corporate asset purchases and sales: Theory and evidence, *Journal of Financial Economics* 87, 471–497.

——— , 2011, Research and development, profits and firm value: A structural estimation, Manuscript, Federal Reserve Board.

Warusawitharna, Missaka, and Toni M. Whited, 2012, Equity market misvaluation and firm financial policies, Manuscript, University of Rochester.

Whited, Toni M., 1992, Debt, liquidity constraints, and corporate investment: Evidence from panel data, *Journal of Finance* pp. 1425–1460.

——— , and Guojun Wu, 2006, Financial constraints risk, *Review of Financial Studies* 19, 531–559.

Wright, Jonathan, 2003, Detecting lack of identification in GMM, *Econometric Theory* 19, 322–330.

Zwiebel, Jeffrey, 1996, Dynamic capital structure under managerial entrenchment, *American Economic Review* pp. 1197–1215.

Table 1: Symbol Definitions

	Symbol	Definition
Section 2		
	X_t	project or firm cash flows
	μ	the instantaneous growth rate of cash flows
	σ	the instantaneous volatility of cash flows
	$dW_t^{\mathbb{Q}}$	a standard Brownian motion process under the risk-neutral measure
	I	the cost (amount) of investment
	r	the risk-free interest rate
	$V(\cdot)$	firm value
	$A(\cdot)$	the date-0 value of an Arrow-Debreu security
	ξ_1, ξ_2	roots of the fundamental quadratic equation
	S_t	the value of equity
	D_t	the value of debt
	X_D	the default threshold
	c	the coupon rate of perpetual debt
	γ	a multiplicative constant defining the default threshold
	α	deadweight default costs
	F	the unlevered firm value
	TB	the present value of the tax benefits of debt
	DC	the present value of default costs
	$L(\cdot)$	the market leverage ratio
	$QML(\cdot)$	quasi-market leverage ratio
	X_R	refinancing threshold
	T_X	time that X is reached for the first time
	$R(\cdot)$	value of the debt payoff at refinancing
	ω	premium associated with the right to recall debt
	q	the proportional cost of issuing debt
	π_t	realized cash flow
	ϵ_t	temporary component of realized cash flow
	λ	intensity of a Poisson process
	$\phi(\cdot)$	Poisson density
	T_s	time of arrival of a temporary shock
	T_r	time of reversal of a temporary shock

Section 3		
	$e(\cdot)$	cash flows to equity holders
	k	capital stock
	z	productivity/demand shock
	$\pi(\cdot)$	profit function
	$\psi(\cdot)$	investment adjustment cost function
	I	investment
	$V(\cdot)$	equity or firm value
	r	risk free interest rate
	$\mathbb{E}_t$	expectations operator conditional on information known at time t
	δ	rate of capital depreciation
	χ	shadow value of capital
	ε	innovation to the $AR(1)$ process for z
	σ_ε	standard deviation of ε
	ρ	serial correlation of the $AR(1)$ process for z
	$g(z' \mid z)$	transition function for z
	θ	profit function curvature
	$h(\cdot)$	policy function
	ψ_0, ψ_1	fixed and quadratic investment adjustment cost parameters
	$\eta(\cdot)$	external financing/equity issuance costs
	η_0, η_1	fixed and linear issuance cost parameters
	p	cash
	b	debt
	s	fraction of debt that can be collateralized
	τ	corporate tax rate on profits and interest
	ζ	Lagrange multiplier
	$\tilde{r}$	interest rate on risky debt
Section 4		
	x	real data vector
	$M(x)$	vector of real data moments
	y	simulated data vector
	$m(y, \beta)$	vector of simulated moments
	β	parameter vector to be estimated
	W	weight matrix
	$Q(x, y, \beta)$	SMM objective function
	N	number of observations in x
	S	ratio of the dimension of y to x
	u_i	random effect
	$f(\cdot)$	generic symbol for a density

Table 2: Comparative Statics of the Dynamic Investment Model

This table presents the comparative statics of the dynamic investment model. To derive the table, we assume that $X_0 < X_I$. For subsequent comparison with the standard NPV criterion, we either assume that $X_I^{NPV} \geq X_0$, where X_I^{NPV} is the cash flow level at which NPV is zero, or that the $+$ and $-$ signs indicate weak (greater or equal, smaller or equal) rather than strong relations.

	Sign of change in variable for an increase in:			
Variable	μ	σ	I	r
X_I	$-$	$+$	$+$	$+$
X_I^{NPV}	$-$	n.a.	$+$	$+$
$X_I - X_I^{NPV}$	$+$	$+$	$+$	$+$
$V(X_0)$	$+$	$+$	$-$	$-$
$V^{NPV}(X_0)$	$+$	n.a.	$-$	$-$
$V(X_0) - V^{NPV}(X_0)$	$-$	$+$	$+$	$+$

Table 3: Comparative Statics of the Optimal Static Capital Structure Model

Variable	Sign of change in variable for an increase in: μ	σ	r	α	τ
Endogenous case					
γ_D	$-$	$-$	$+$	n.a.	n.a.
c	$+$.	$-/+$	$-$	$-$	$+$
L	$+$	$-$	$+$	$-$	$+$
X_D	$+$	$-$	$+$	$-$	$+$
$\frac{V(X_0)-F(X_0)}{F(X_0)}$	$+$	$-$	$+$	$-$	$+$
Exogenous case					
c	$-/+$	$-$	$+$	$-$	$+$
L	$-$	$-$	$+$	$-$	$+$
X_D	$-/+$	$-$	$+$	$-$	$+$
$\frac{V(X_0)-F(X_0)}{F(X_0)}$	$-$	$-$	$+$	$-$	$+$

Table 4: Static Capital Structure

This table reports the optimal coupon, c, date-0 leverage ratio, L, and the difference between levered and unlevered firm values for the optimal static capital structure model. The benchmark set of parameters is: $r = 0.05$, $\mu = 0.02$, $\sigma = 0.25$, $\tau = 0.2$, and $\alpha = 0.1$. The first three columns are for the exogenous default case, and the last three columns are for the endogenous default case.

	c	L	$\frac{V-F}{F}\%$	c	L	$\frac{V-F}{F}\%$
Benchmark	0.32	0.25	0.03	1.36	0.70	0.11
$\sigma = 0.15$	0.50	0.37	0.05	1.28	0.75	0.14
$\sigma = 0.40$	0.16	0.13	0.01	1.73	0.66	0.09
$\tau = 0.35$	0.40	0.34	0.07	1.49	0.77	0.25
$\tau = 0.10$	0.22	0.16	0.01	1.14	0.59	0.04
$\alpha = 0.5$	0.13	0.10	0.01	0.84	0.48	0.07
$\alpha = 0.05$	0.39	0.31	0.03	1.48	0.74	0.12

Table 5: Dynamic Capital Structure

This table reports the optimal coupon, c, date-0 leverage ratio, L, and the difference between levered and unlevered firm values for the optimal dynamic and static capital structure models. The benchmark set of parameters is: $r = 0.05$, $\mu = 0.02$, $\sigma = 0.25$, $\tau = 0.2$, and $\alpha = 0.1$. The first three columns are for the endogenous default case of the static capital structure model and the last three columns are for the dynamic capital structure model.

	Static			Dynamic		
	c	L	$\frac{V-F}{F}\%$	c	L	$\frac{V-F}{F}\%$
Benchmark	1.33	0.69	0.10	1.12	0.54	0.18
$\sigma = 0.15$	1.26	0.75	0.13	1.13	0.63	0.20
$\sigma = 0.40$	1.67	0.65	0.08	1.24	0.45	0.19
$\tau = 0.35$	1.48	0.77	0.24	1.32	0.57	0.51
$\tau = 0.10$	1.08	0.57	0.03	0.88	0.46	0.06
$\alpha = 0.5$	0.82	0.48	0.06	0.67	0.37	0.11
$\alpha = 0.05$	1.45	0.73	0.11	1.22	0.57	0.19
$q = 0.005$	1.35	0.69	0.10	1.06	0.51	0.19
$q = 0.05$	1.19	0.66	0.07	1.06	0.57	0.11

Table 6: True Dynamics of Leverage in Dynamic Capital Structure

This table reports the true dynamics of leverage in the dynamic capital structure model. The benchmark set of parameters is: $r = 0.05$, $\mu = 0.02$, $\sigma = 0.25$, $\tau = 0.2$, and $\alpha = 0.1$. The first three columns are for the exogenous default case, and the last three columns are for the endogenous default case.

	Static L	$L(X_0)$	mean L_{TD}	st. dev. L_{TD}	min L_{TD}
Benchmark	0.70	0.54	0.66	0.17	0.35
$\sigma = 0.15$	0.76	0.63	0.65	0.12	0.44
$\sigma = 0.40$	0.66	0.45	0.68	0.20	0.28
$\tau = 0.35$	0.77	0.56	0.70	0.15	0.42
$\tau = 0.10$	0.60	0.46	0.59	0.18	0.27
$\alpha = 0.5$	0.49	0.37	0.54	0.18	0.27
$\alpha = 0.05$	0.74	0.57	0.68	0.16	0.37
$q = 0.005$	0.70	0.51	0.66	0.16	0.37
$q = 0.05$	0.72	0.57	0.59	0.20	0.23

Table 7: Cross-Sectional regressions of Leverage on Profitability

This table reports the results of cross-sectional regressions on the level of the quasi-market leverage ratio, QML, on profitability and control variables. The results are from Strebulaev (2007).

	$QML(X_0)$	QML_{TD}	QML_{TD}, 10%	QML_{TD}, 90%
Constant	0.24	0.62	0.55	0.71
Profitability	5.88	-0.78	-1.53	-0.22
Control variables	Yes	Yes	NaN	NaN
R^2	0.89	0.08	0.06	0.10

Table 8: Capital Structure with Strategic Renegotiation

This table reports the optimal bank coupon, c_B, optimal public debt coupon, c, date-0 leverage ratio, L, and the difference between levered and unlevered firm value for the optimal static capital structure model with only public debt, only bank debt, and both bank and public debt. The benchmark set of parameters is: $r = 0.05$, $\mu = 0.02$, $\sigma = 0.25$, $\tau = 0.2$, and $\alpha = 0.1$. The first three columns are for the case of only public debt; the next three columns are for case of only bank debt; and the last three columns are for the case of both bank and public debt.

	Only public debt			Only bank debt			Both bank and public debt			
	c	L	$(V-F)/F$	c_B	L	$(V-F)/F$	c	c_B	L	$(V-F)/F$
Benchmark	1.36	0.70	0.11	2.29	0.76	0.18	0.16	2.29	0.79	0.19
$\mu = 0.005$	0.94	0.69	0.10	1.68	0.76	0.18	0.10	1.68	0.78	0.19
$\mu = 0.035$	2.65	0.71	0.11	4.21	0.76	0.18	0.35	4.21	0.80	0.20
$\sigma = 0.15$	1.28	0.75	0.14	1.67	0.76	0.18	0.23	1.67	0.83	0.20
$\sigma = 0.40$	1.73	0.66	0.09	3.60	0.76	0.18	0.11	3.60	0.77	0.19
$\tau = 0.35$	1.49	0.77	0.25	1.86	0.68	0.32	0.36	1.86	0.65	0.38
$\tau = 0.10$	1.14	0.59	0.04	2.58	0.83	0.09	0.06	2.58	0.85	0.09
$\alpha = 0.5$	0.84	0.48	0.07	1.27	0.45	0.10	0.35	1.27	0.53	0.13
$\alpha = 0.05$	1.48	0.74	0.12	2.42	0.80	0.19	0.14	2.42	0.82	0.20

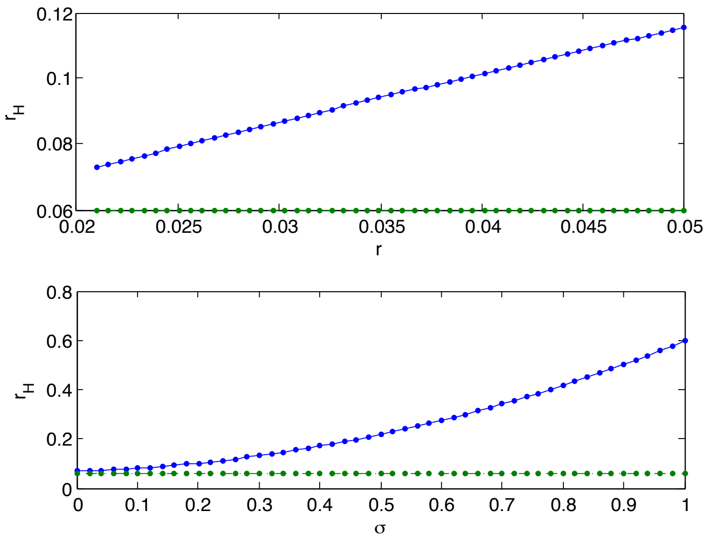

Figure 1: **Real options investment model. Comparative statics of the hurdle rate, r_H.** This figure shows the comparative statics of the hurdle rate, r_H, with respect to the interest rate, r, and asset volatility, σ, for both the NPV model (smaller values) and the real options model.

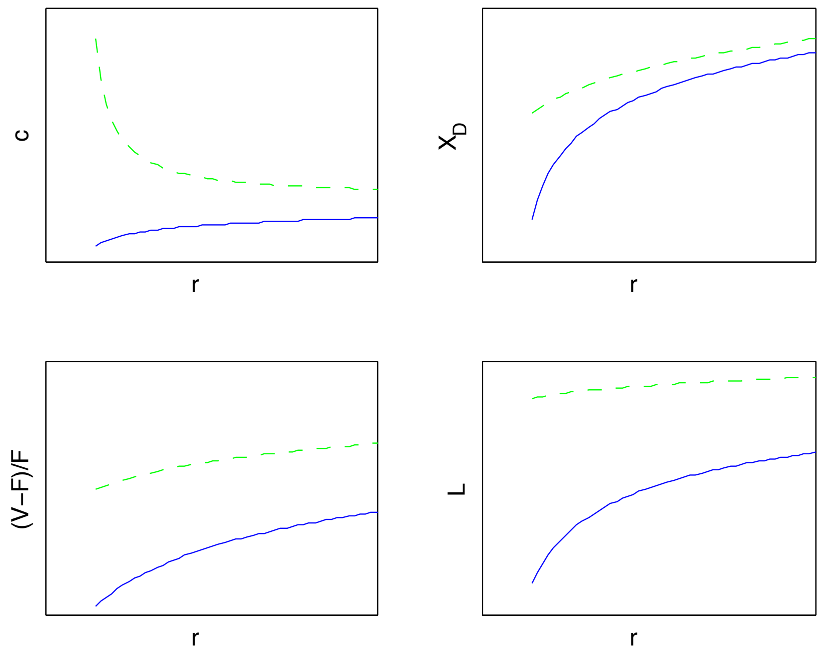

Figure 2: **Optimal static capital structure model. Comparative statics with respect to the risk-free interest rate.** This figure shows the comparative statics of the optimal coupon rate, c, the default boundary, X_D, optimal leverage, L, and the difference in values between levered and unlevered firms, $\frac{V-F}{F}$, all with respect to the risk-free interest rate, r, in the optimal static capital structure model. The dashed lines show the endogenous default case ($\gamma = \frac{\xi_1}{\xi_1 - 1}\frac{r-\mu}{r}$), and the solid lines show the exogenous default case ($\gamma = 1$).

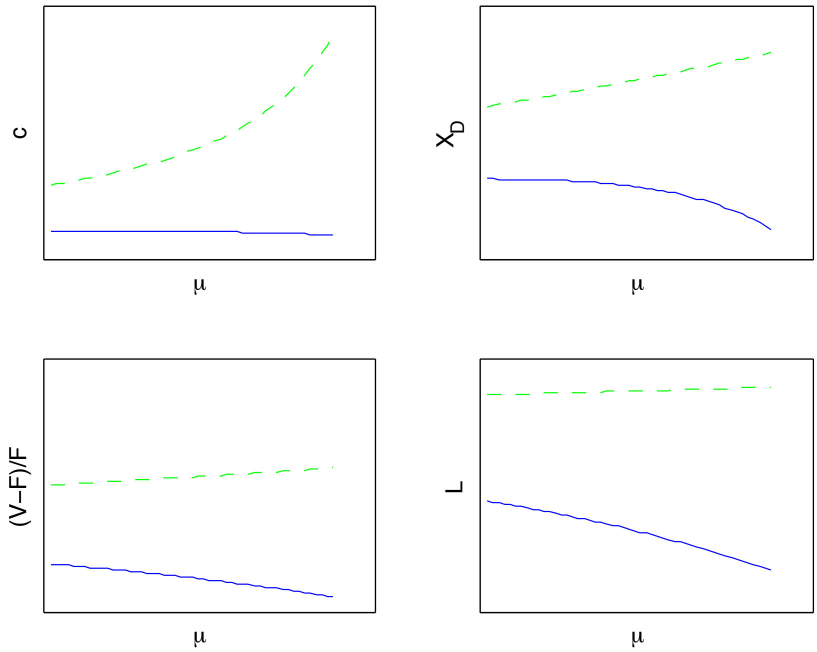

Figure 3: **Optimal static capital structure model. Comparative statics with respect to the growth rate.** This figure shows the comparative statics of the optimal coupon rate, c, the default boundary, X_D, optimal leverage, L, and the difference in values between levered and unlevered firms, $\frac{V-F}{F}$, all with respect to the growth rate, μ, in the optimal static capital structure model. The dashed lines show the endogenous default case ($\gamma = \frac{\xi_1}{\xi_1 - 1} \frac{r-\mu}{r}$), and the solid lines show the exogenous default case ($\gamma = 1$).

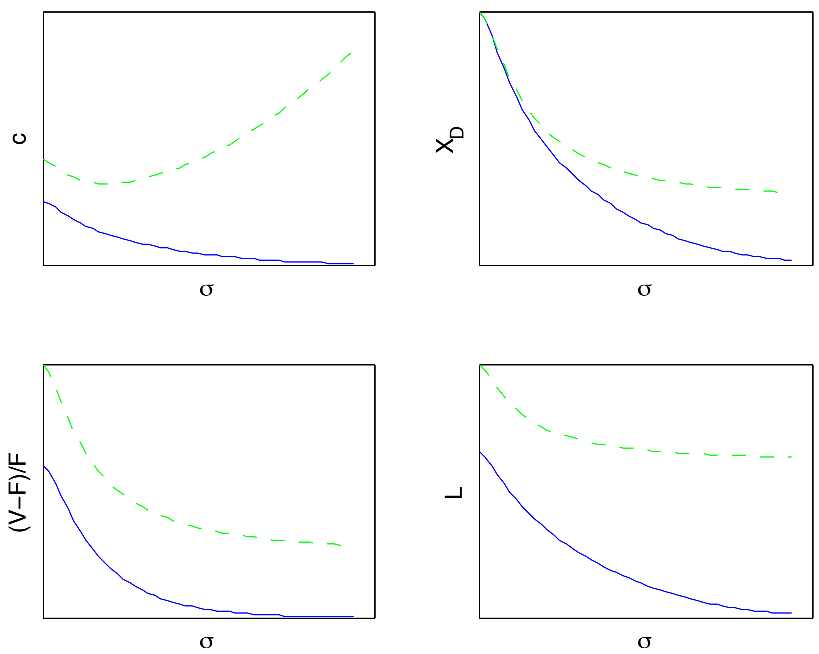

Figure 4: **Optimal static capital structure model. Comparative statics with respect to asset volatility.** This figure shows the comparative statics of the optimal coupon rate, c, the default boundary, X_D, optimal leverage, L, and the difference in values between levered and unlevered firms, $\frac{V-F}{F}$, all with respect to asset volatility, σ, in the optimal static capital structure model. The dashed lines show the endogenous default case ($\gamma = \frac{\xi_1}{\xi_1 - 1}\frac{r-\mu}{r}$), and the solid lines show the exogenous default case ($\gamma = 1$).

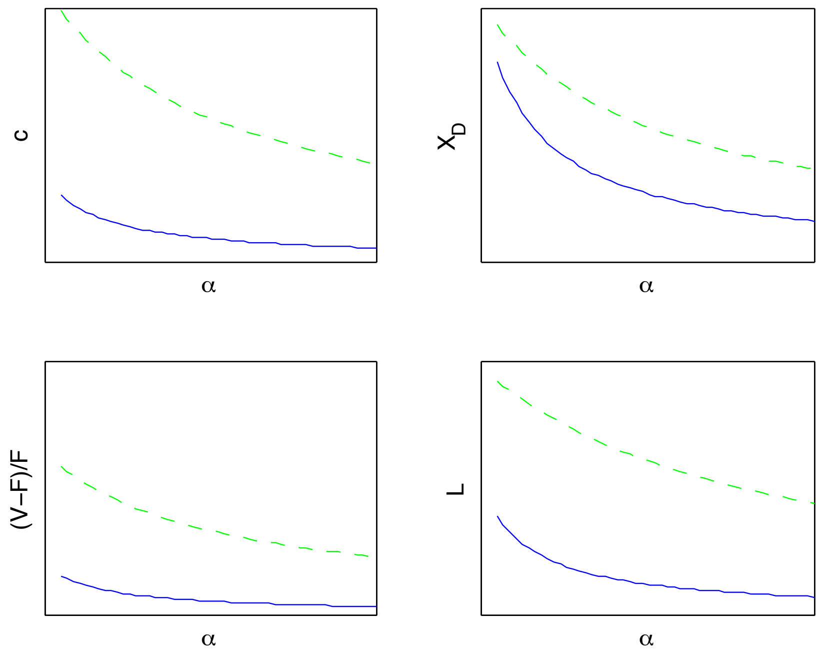

Figure 5: **Optimal static capital structure model. Comparative statics with respect to bankruptcy costs.** This figure shows the comparative statics of the optimal coupon rate, c, the default boundary, X_D, optimal leverage, L, and the difference in values between levered and unlevered firms, $\frac{V-F}{F}$, all with respect to the bankruptcy cost parameter, α, in the optimal static capital structure model. The dashed lines show the endogenous default case ($\gamma = \frac{\xi_1}{\xi_1 - 1}\frac{r-\mu}{r}$), and the solid lines show the exogenous default case ($\gamma = 1$).

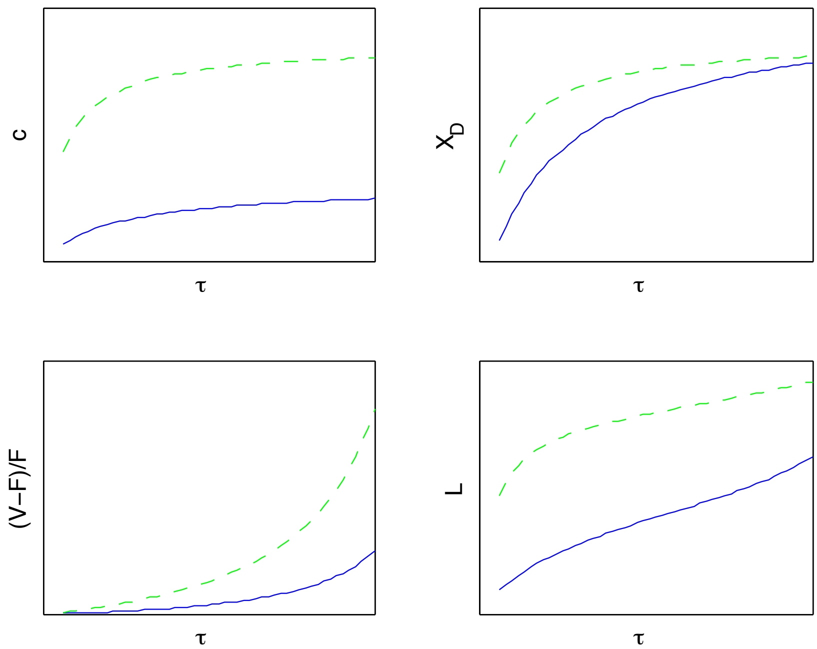

Figure 6: **Optimal static capital structure model. Comparative statics with respect to corporate tax rate.** This figure shows the comparative statics of the optimal coupon rate, c, the default boundary, X_D, optimal leverage, L, and the difference in values between levered and unlevered firms, $\frac{V-F}{F}$, all with respect to the corporate income tax rate, τ, in the optimal static capital structure model. The dashed lines show the endogenous default case ($\gamma = \frac{\xi_1}{\xi_1 - 1}\frac{r-\mu}{r}$), and the solid lines show the exogenous default case ($\gamma = 1$).

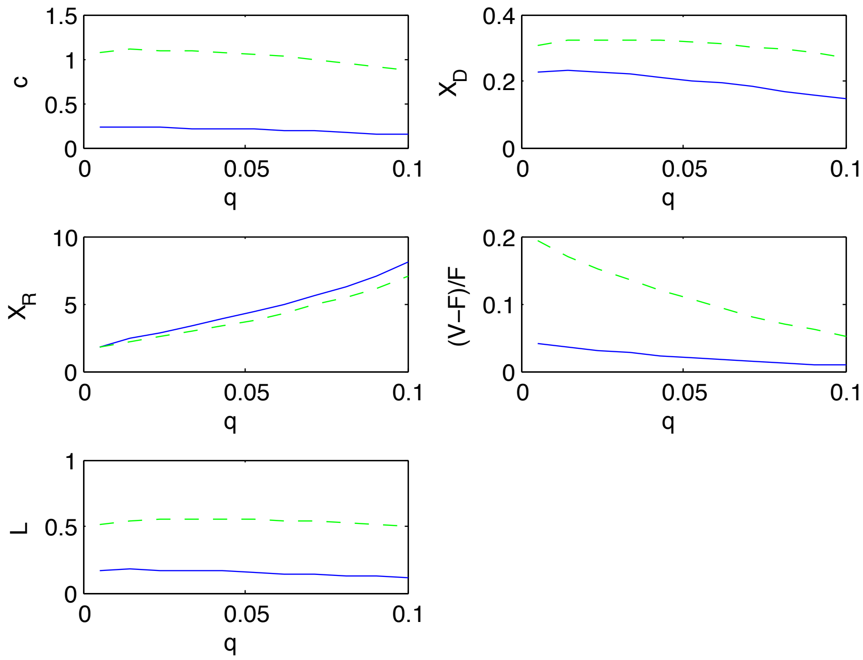

Figure 7: **Optimal dynamic capital structure model. Comparative statics with respect to debt issuance cost.** This figure shows the comparative statics of the optimal coupon, c, the default boundary, X_D, the refinancing boundary, X_R, optimal leverage, L, and the difference in values between levered and unlevered firms, $\frac{V-F}{F}$, all with respect to the debt issuance cost parameter, q, in the optimal dynamic capital structure models. The dashed lines show the endogenous default case, and the solid lines show the exogenous default case.

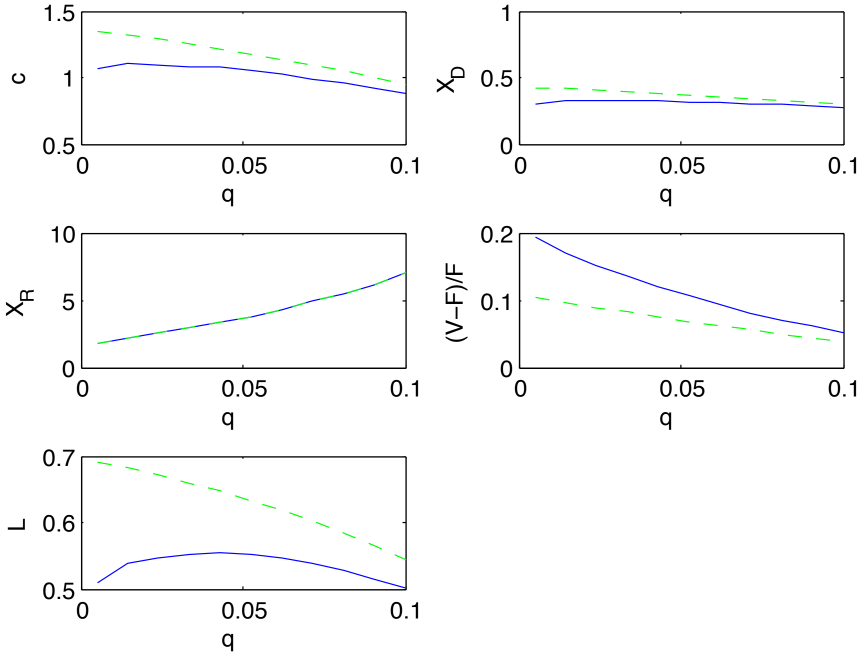

Figure 8: **Optimal dynamic and static capital structure models. Comparative statics with respect to debt issuance cost.** This figure shows the comparative statics of the optimal coupon, c, the default boundary, X_D, optimal leverage, L, and the difference in values between levered and unlevered firms, $\frac{V-F}{F}$, all with respect to the debt issuance cost parameter, q, in the optimal dynamic and static capital structure models. For the dynamic model, the results on refinancing boundary, X_R, are also shown. The dashed lines show the static capital structure model (with debt issuance costs), and the solid lines show the dynamic capital structure model.

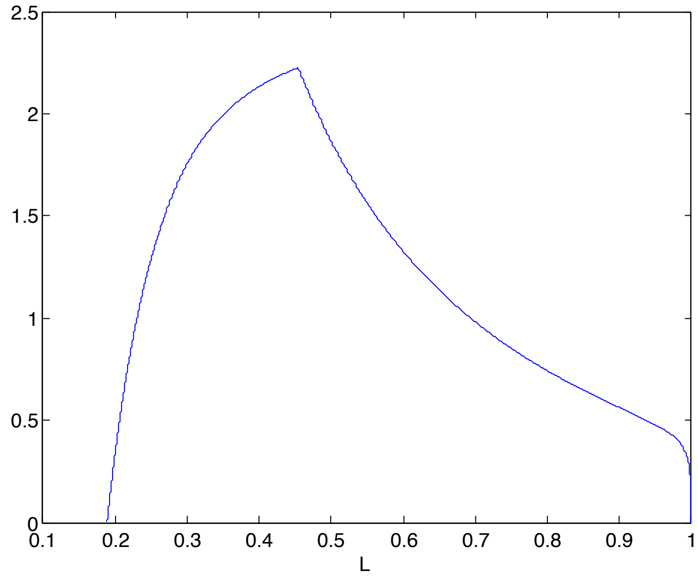

Figure 9: **Dynamic capital structure model. The cross-sectional distribution of the leverage ratio in true dynamics.** The figure shows the cross-sectional distribution of the leverage ratio, L, in true dynamics in the dynamic capital structure model.

Figure 10: Policy Functions for Investment Models

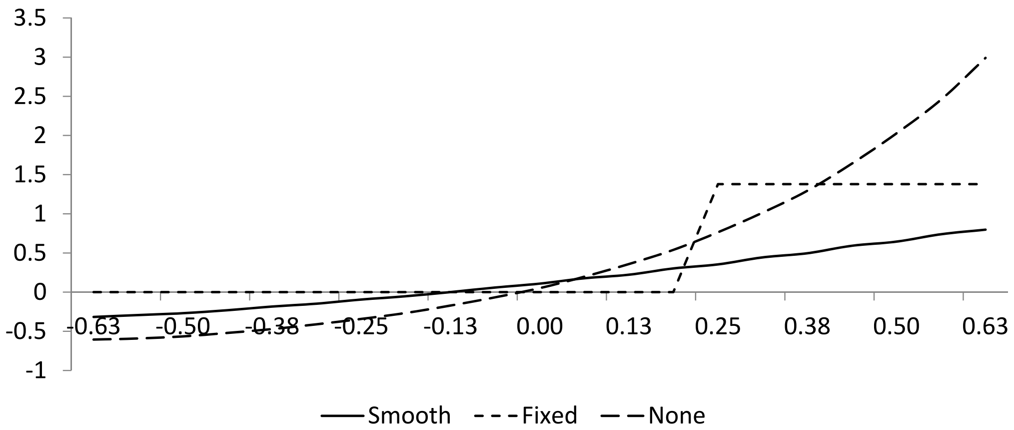

Figure 10 depicts three policy functions for optimal investment. These functions map the current productivity shock and the steady state capital stock into optimal future investment. The plots are evaluated at the steady state capital stock, the log productivity shock is on the horizontal axis, and the rate of investment is on the vertical axis. Each policy function is for a version of the simple investment model from Section 3.1. One version contains no adjustment costs, one contains convex, smooth adjustment costs, and one contains fixed adjustment costs.

Figure 11: Comparative Statics: Investment Model with Adjustment Costs

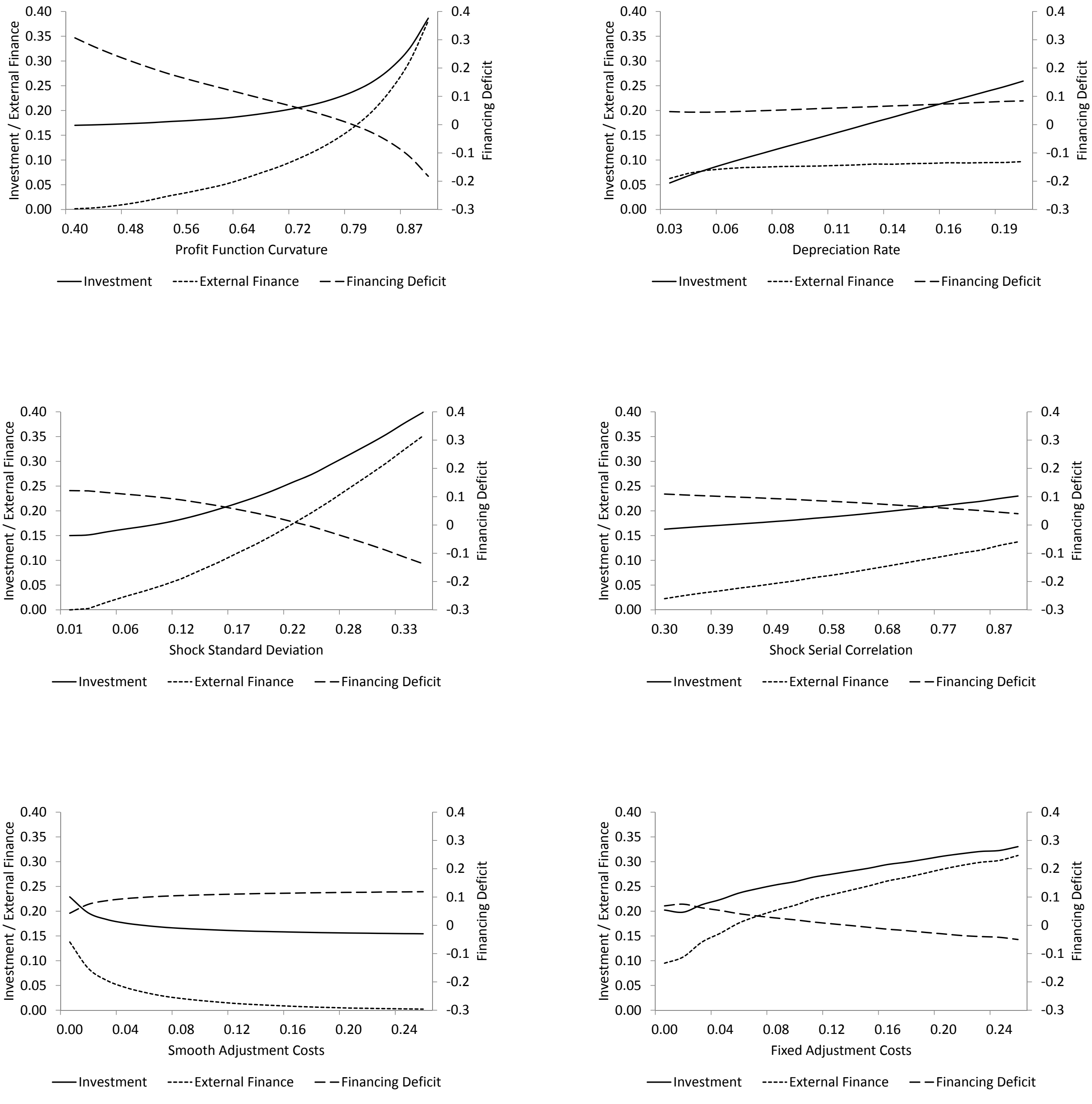

Figure 11 depicts average investment, external financing, and financing deficits as a function of six model parameters: profit function curvature, the depreciation rate of capital, the standard deviation of the profit shock, the serial correlation of the profit shock, a smooth adjustment cost parameter, and a fixed adjustment cost parameter. All quantities are scaled by total firm assets. External financing is defined to be zero if the firm distributes funds to shareholders and positive otherwise. The financing deficit is defined as internal funds minus investment.

Figure 12: Investment-Cash Flow Sensitivity in a Dynamic Investment Model

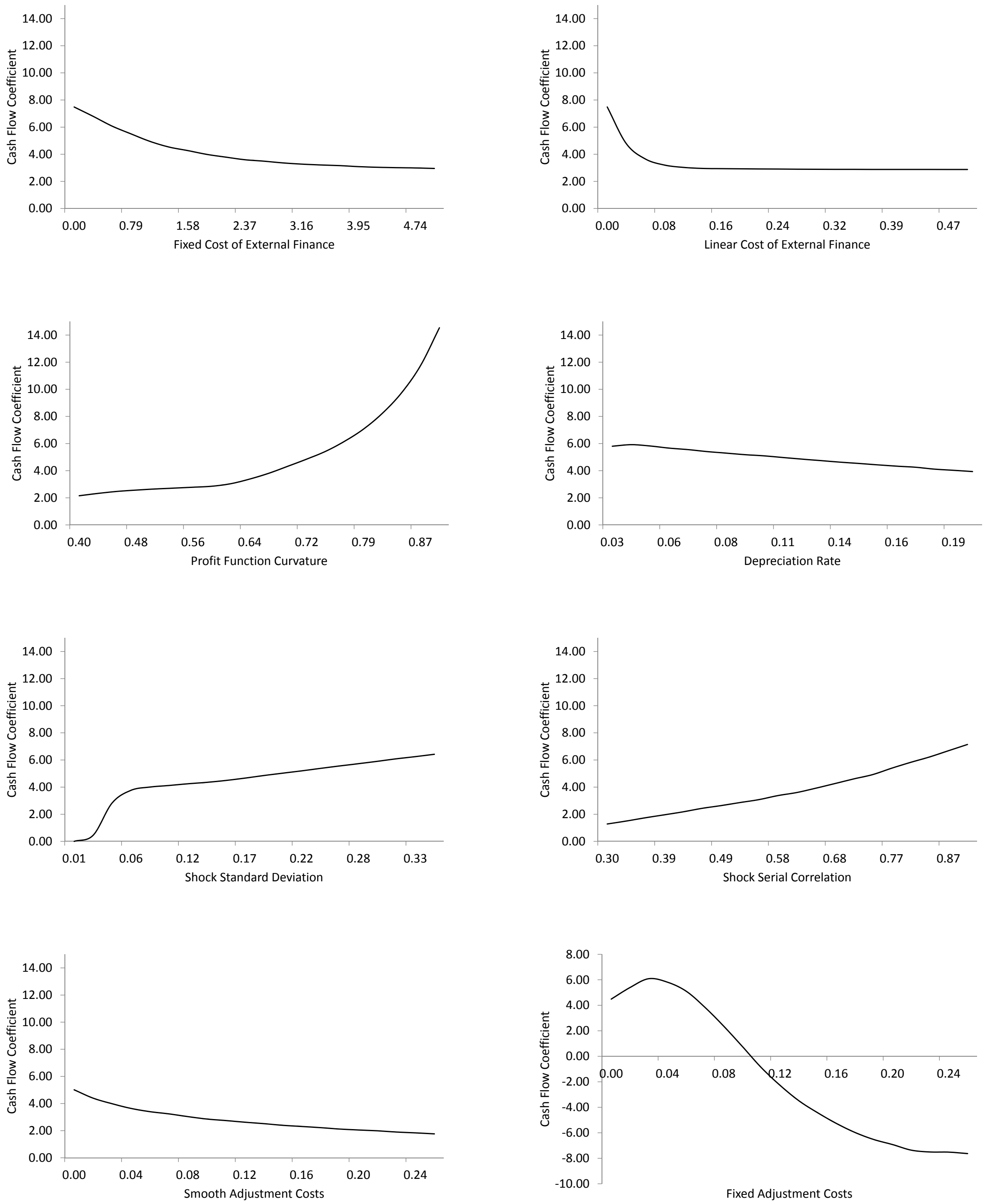

Figure 12 depicts the coefficient on cash flow in a regression of investment on Tobin's q, and cash flow, using simulated data. This coefficient is plotted as a function of the fixed and linear costs of external finance, as well as a function of profit function curvature, the depreciation rate of capital, the standard deviation of the profit shock, the serial correlation of the profit shock, a smooth adjustment cost parameter, and a fixed adjustment cost parameter.

Figure 13: Policy Functions for a Dynamic Cash Model

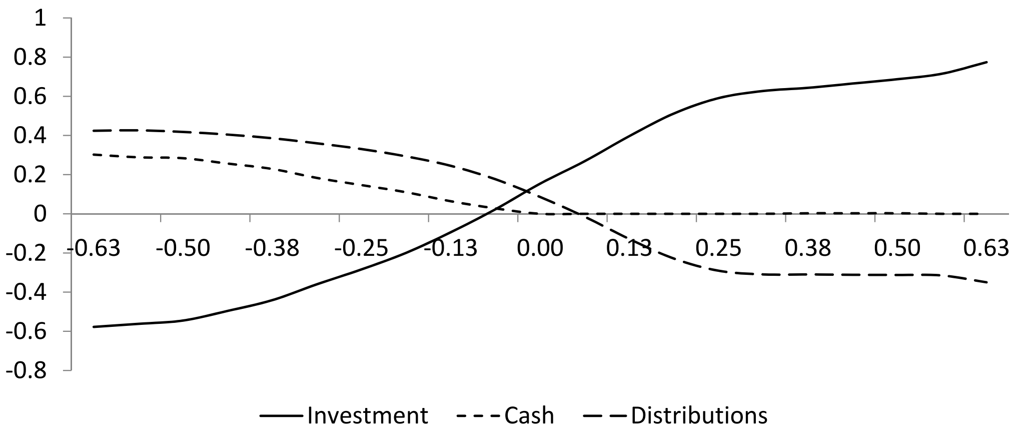

Figure 13 depicts the policy functions for the model from Section 3.3 These functions map the current log productivity shock and the steady state capital stock into optimal future investment, optimal future cash balances, and optimal future distributions to shareholders. The plots are evaluated at the steady state capital stock, and the log productivity shock is on the horizontal axis. Each variable is scaled by the capital stock, and negative distributions are equivalent to equity issuances.

Figure 14: Average Cash Balances

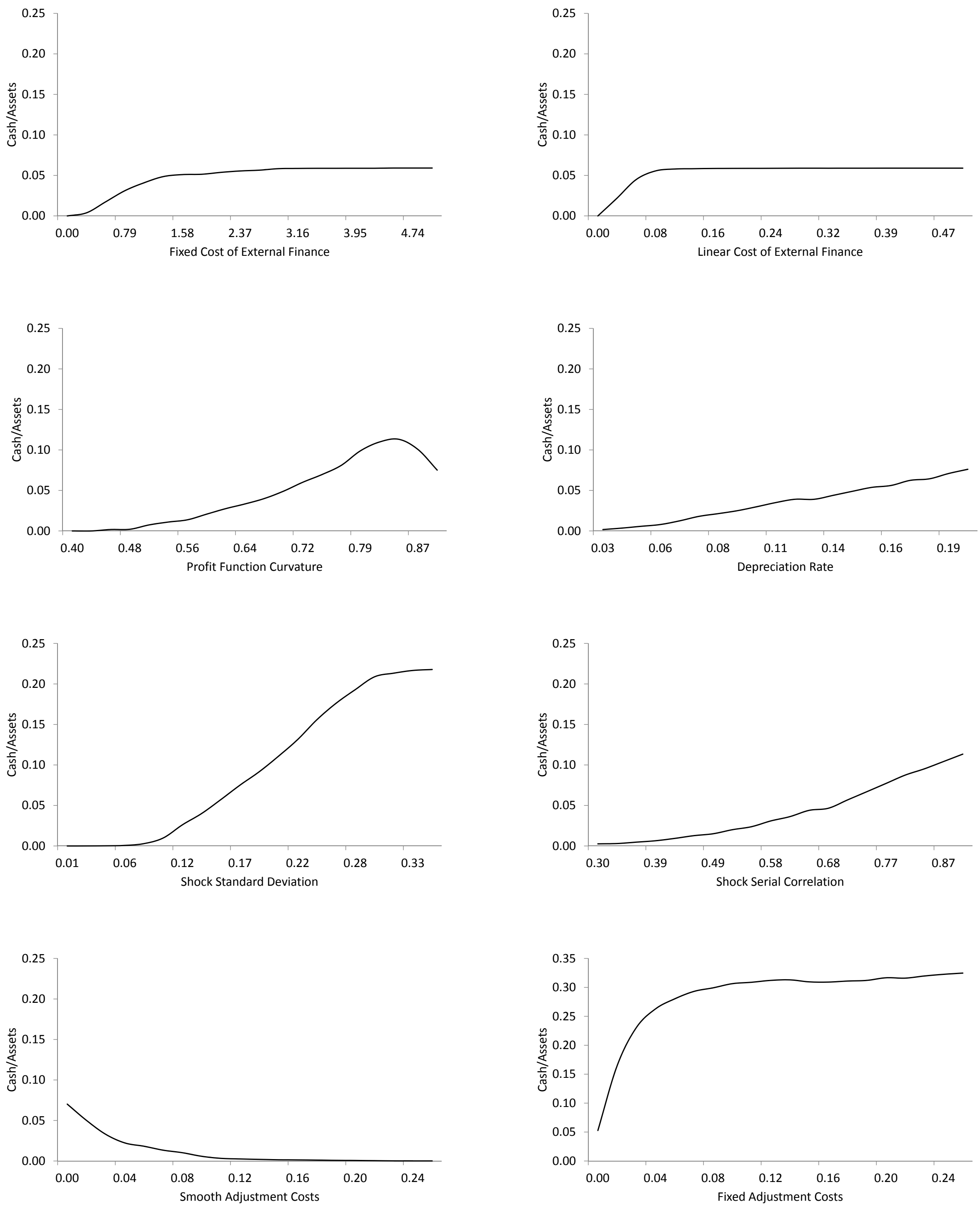

Figure 14 depicts the average ratio of cash to assets. This ratio is plotted as a function of the fixed and linear costs of external finance, as well as a function of profit function curvature, the depreciation rate of capital, the standard deviation of the profit shock, the serial correlation of the profit shock, a smooth adjustment cost parameter, and a fixed adjustment cost parameter.

Figure 15: Optimal Debt Policy: Financing Surplus

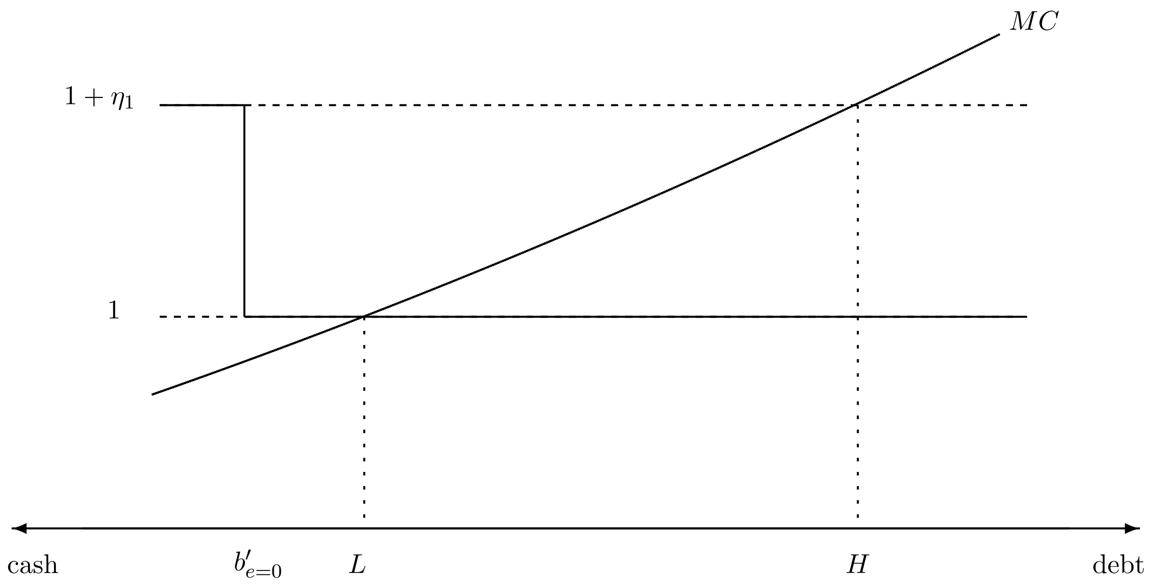

Figure 15 depicts the first-order conditions for optimality from the model in Section 3.4. The upward sloping line is the marginal cost (MC) of debt financing. The solid portion of the two horizontal lines is the marginal benefit schedule. $b'_{e=0}$ is the level of debt at which the sources of funds equals the uses of funds. Optimal leverage can be found at the point L.

Figure 16: Optimal Debt Policy: Financing Deficit

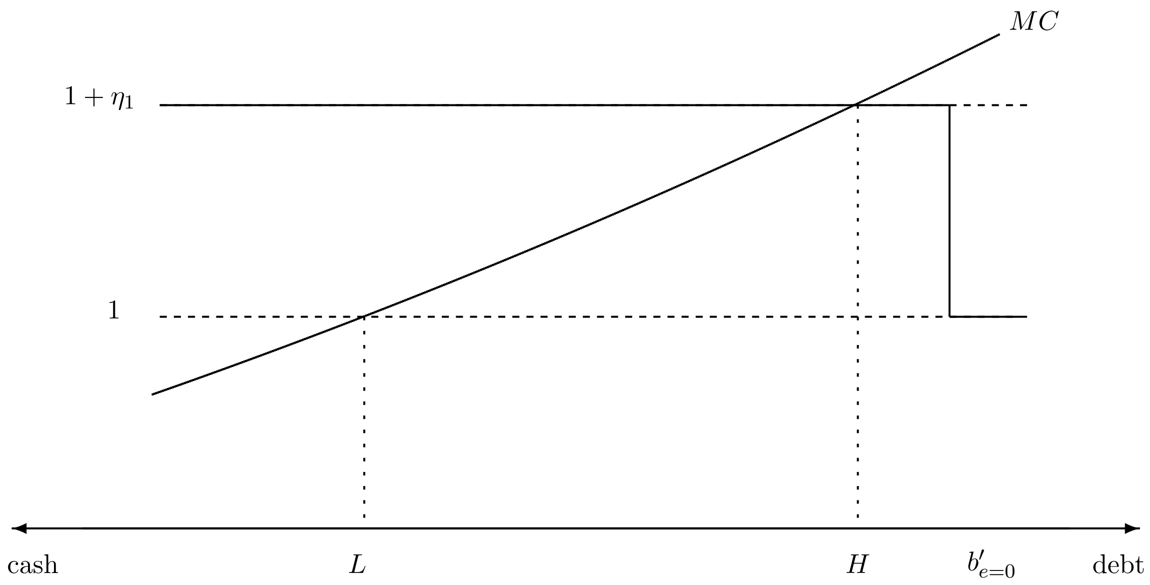

Figure 16 depicts the first-order conditions for optimality from the model in Section 3.4. The upward sloping line is the marginal cost (MC) of debt financing. The solid portion of the two horizontal lines is the marginal benefit schedule. $b'_{e=0}$ is the level of debt at which the sources of funds equals the uses of funds. Optimal leverage can be found at the point H.

Figure 17: Policy Functions for a Dynamic Debt Model

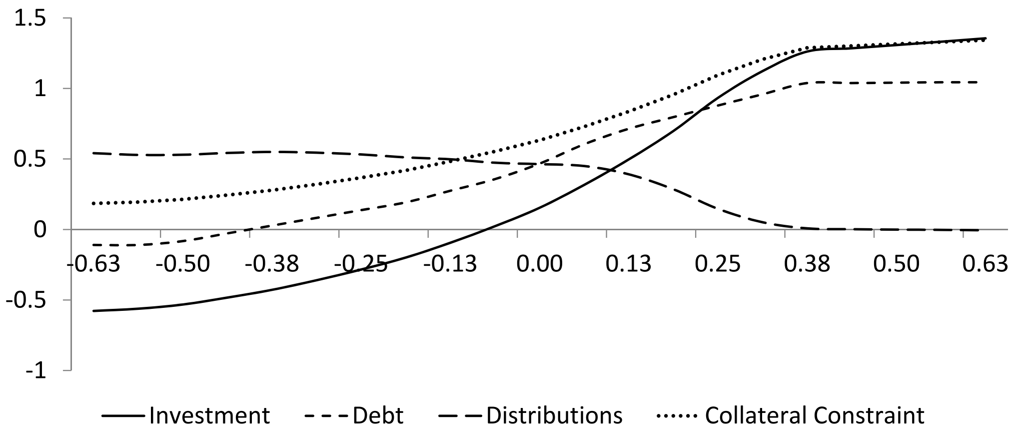

Figure 17 depicts the policy functions for the model from Section 3.4 These functions map the current log productivity shock and the steady state capital stock into optimal future investment, optimal future debt, and optimal future distributions to shareholders. The plots are evaluated at the steady state capital stock, and the log productivity shock is on the horizontal axis. Each variable is scaled by the capital stock, and negative distributions are equivalent to equity issuances.

Figure 18: Average Leverage and Debt Capacity

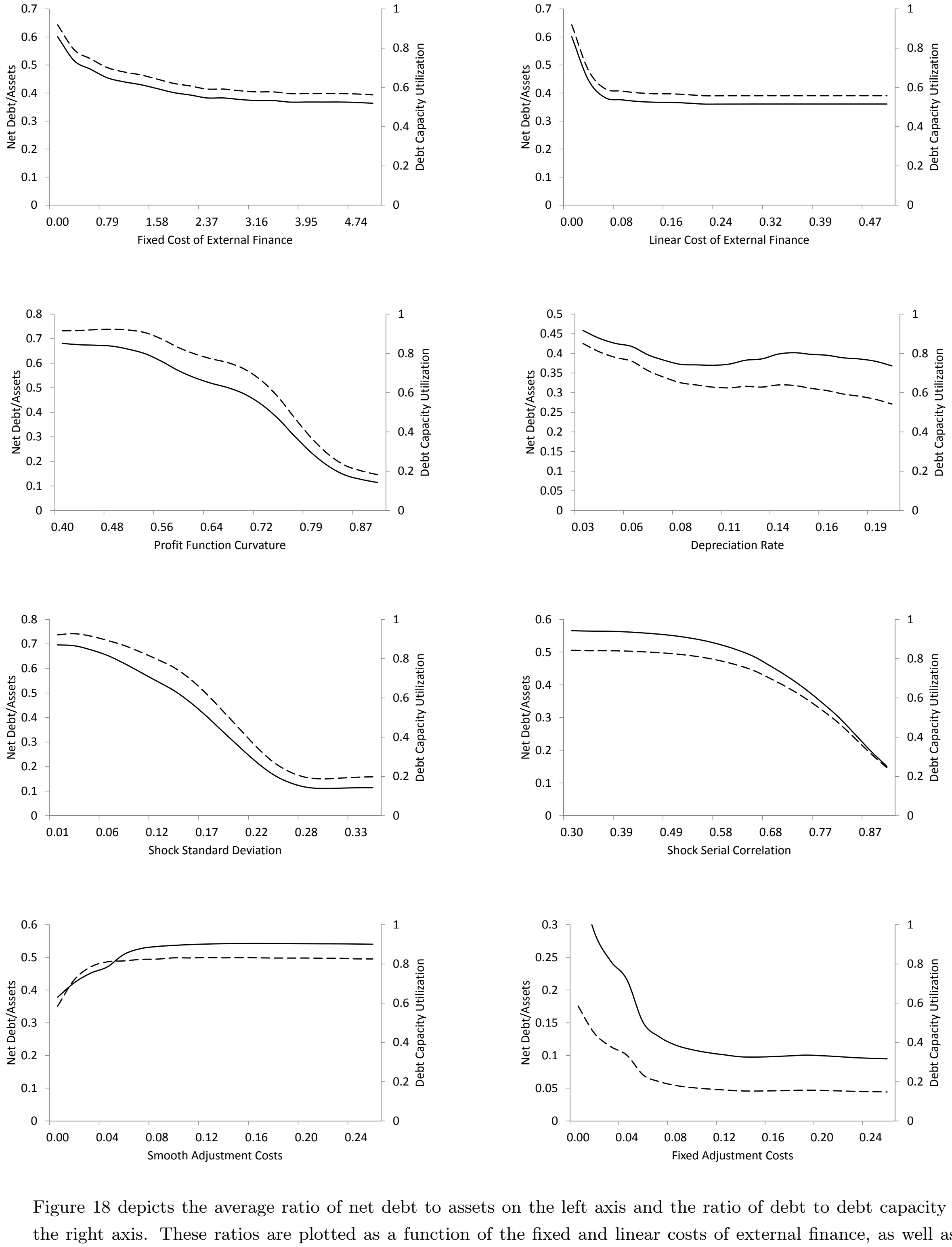

Figure 18 depicts the average ratio of net debt to assets on the left axis and the ratio of debt to debt capacity on the right axis. These ratios are plotted as a function of the fixed and linear costs of external finance, as well as a function of profit function curvature, the depreciation rate of capital, the standard deviation of the profit shock, the serial correlation of the profit shock, a smooth adjustment cost parameter, and a fixed adjustment cost parameter.