ANNUAL REVIEWS

Annual Review of Financial Economics

The Marginal Value of Cash: Corporate Savings, Investment, and Financing

Patrick Bolton,[1,4,5] Hui Chen,[2,5] and Neng Wang[3,5,6]

[1]Imperial College London, London, United Kingdom

[2]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA; email: huichen@mit.edu

[3]Cheung Kong Graduate School of Business, Beijing, China

[4]Centre for Economic Policy Research, London, United Kingdom

[5]National Bureau of Economic Research, Cambridge, Massachusetts, USA

[6]Asian Bureau of Finance and Economic Research, Singapore

Annu. Rev. Financ. Econ. 2024. 16:295–324

First published as a Review in Advance on September 24, 2024

The *Annual Review of Financial Economics* is online at financial.annualreviews.org

https://doi.org/10.1146/annurev-financial-111021-103759

JEL codes: E22, E44, G3

Keywords

marginal value of cash, investment, financing, q theory, liquidity

Abstract

The fact that internal liquidity is a key source of corporate funding puts the marginal value of cash at the center of a variety of firm decisions, including investment, payout, financing, savings, and risk management policies. The marginal value of cash is inherently a dynamic concept, because a firm facing financing frictions has to be forward-looking, managing its asset and liability structures in a unified framework and carefully trading off the use of liquidity across time and states. We present a dynamic framework for corporate liquidity management and survey the related literature, with a focus on the determinants of the marginal value of cash and its ubiquitous role in firm decisions.

1. INTRODUCTION

It is barely an exaggeration to say that when it comes to the question of how corporations should finance their investments, corporate finance textbooks have devoted an almost exclusive attention to one margin: the optimal choice of debt and equity. The simplified representation of a balance sheet in textbooks is then as shown in **Figure 1**, with the asset side composed of the value of operating assets and the liability side by debt and equity.

In practice, however, chief financial officers (CFOs) face at least one additional key margin: the optimal amount of corporate savings (cash and marketable securities) to accumulate and hold along with operating assets on the asset side of the balance sheet, together with the optimal choice of financing between internal and external funds. A typical balance sheet a CFO looks at is then as shown in **Figure 2**.

The reason why CFOs need to take account of this second margin has to do with the costs involved in raising outside financing. Numerous empirical studies have documented that the costs of raising outside debt and equity can be significant and most firms' capacity to borrow is limited. Accordingly, to avoid incurring external financing costs, firms seek to accumulate a cushion of internal funds by saving retained earnings and to fund their capital expenditures with internal funds to the extent that this is feasible.

Faced with the two margins, the CFO's problem is naturally much more complex. The two margins interact in subtle ways, both over time and across different states. The additional complexity involved is one reason why the second margin is typically not included in textbook treatments of capital structure and corporate financing choices for investment. Another reason is that corporate savings were long thought to be relatively small and of second-order relevance. But this has changed dramatically in the past half century, as **Figure 3** illustrates. In 1970, the median cash-to-total assets ratio across all nonfinancial public firms in the United States was 5.4%. It then rose steadily over time and reached 16.4% in 2022. The rise in cash holding among the cash-rich firms is even more striking, with the cash-to-total assets ratio for the 75th percentile rising from 10% in 1970 to 48.8% in 2022.[1]

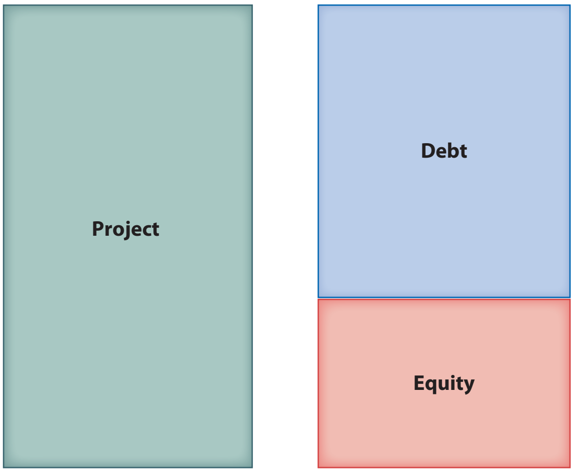

Figure 1

A balance sheet that focuses on debt versus equity. This figure shows a simplified balance sheet that focuses on debt and equity as alternative sources of financing. On the asset side, it does not distinguish between liquid and illiquid assets.

[1]For a survey on the evidence of historical trends and cross-sectional patterns in corporate cash holdings, see also Denis & Wang (2024).

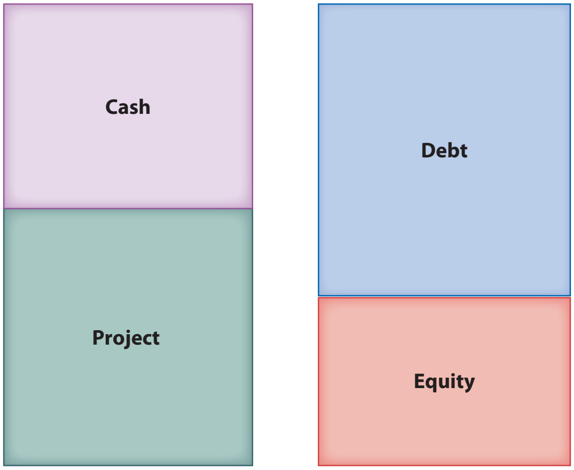

Figure 2

An expanded balance sheet that highlights the role of cash. In this figure, the asset side of the balance sheet is further divided into liquid assets (cash and marketable securities) and illiquid (operating) assets.

A third reason why internally accumulated funds have been overlooked is that, until recently, there were no tractable analytical frameworks available to tackle the analysis of both margins in an integrated way. In this article, we outline such an analytical framework, beginning with a static model to illustrate part of the key intuition and then extending it to a dynamic model, which generates predictions that are easier to be taken to the data. This analytical framework builds on multiple theoretical contributions in the literature, in particular that by Bolton, Chen & Wang (2011).

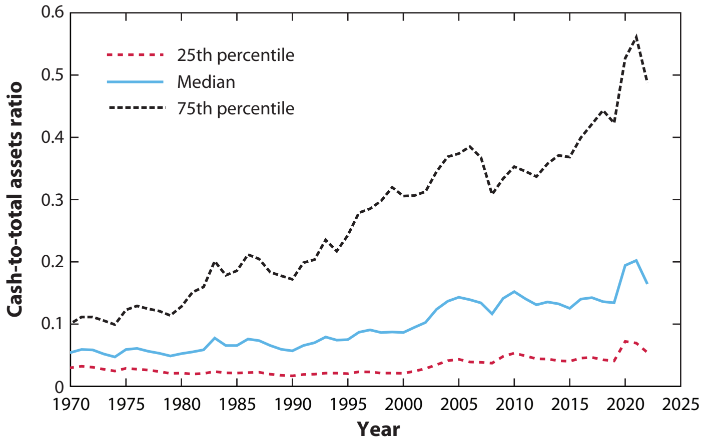

Figure 3

Cash holdings across US nonfinancial public firms. For each year from 1970 to 2022, this figure plots the 25th, 50th, and 75th percentiles of the cross-sectional distribution of cash-to-total assets ratios across US nonfinancial public firms. Data from COMPUSTAT.

There is extensive research on corporate liquidity management, starting with the classic works by Keynes (1936), Baumol (1952), Tobin (1956), and Miller & Orr (1966). Much of the existing theoretical work focuses on static settings, which are covered in the survey by Almeida et al. (2014). In this article, we examine corporate liquidity management from a dynamic perspective, with an emphasis on the determinants of the marginal value of cash and its ubiquitous effects in corporate decisions.

Since internal cash is often the marginal source of funding for investment, wages, risk management, payout, and wages, among other activities, these decisions all depend on a key quantity, the marginal value of cash, that is, how much an extra dollar of internal cash is worth to the firm. This value can be higher than that of a dollar outside the firm due to financing frictions. In this sense, a firm is considered financially constrained when its marginal value of cash is above one; otherwise, the firm is financially unconstrained.

Consider, for example, the decision to make an extra dollar of investment that is funded by internal cash. Under optimality, the marginal q, i.e., the marginal value of investing, has to match the marginal cost of investing, which reflects not only the standard (physical) cost of investment but also the fact that these costs drain internal liquidity:

$$\text{marginal } q = \text{marginal (physical) cost of investing} \times \text{marginal value of cash}. \qquad 1.$$

Next, consider the risk management decision. When a firm tries to reduce its cash flow risks through hedging (using financial instruments such as futures), even if there is no upfront cost for hedging, the margin requirement will be a drain on internal liquidity. This shadow cost is reflected in the cost-benefit analysis for hedging and reduces the firm's hedging demand:

$$\text{marginal benefit of hedging} = \text{cost of margin requirement} \times \text{marginal value of cash}. \qquad 2.$$

A similar argument applies when a firm chooses its financial leverage. The marginal net benefit of debt will again depend on the marginal value of cash. Consider a firm with an extra dollar of income. It can be used in one of three ways: (*a*) paid out to debt-holders as interest (and shielded from corporate income tax), (*b*) paid out to equity-holders as dividends, or (*c*) retained inside the firm. Whereas the famous Miller formula for the net tax benefit of debt focuses on the trade-off between (*a*) and (*b*), the relevant trade-off for a financially constrained firm is mostly between (*a*) and (*c*). This latter comparison leads to the following formula for the net tax benefit of debt:

$$\text{marginal net tax benefit of debt} = \frac{(1-\tau_i) - \text{marginal value of cash} \times (1-\tau_c)}{(1-\tau_i)}, \qquad 3.$$

where τ_c and τ_i denote the marginal tax rate on corporate income and personal interest income, respectively. In the presence of financial constraints, the marginal value of cash to equity-holders is equal to $(1-\tau_e)$ only when the firm is unconstrained: It is indifferent between keeping an extra dollar of cash inside the firm or paying it out as dividend. Only in this special case, we recover the Miller formula. When the financial constraint is severe, the marginal net tax benefit of debt can be negative.

Finally, the marginal value of cash also plays an important role in determining a firm's beta and expected rate of returns. Firm beta depends on the sensitivity of firm value to market-wide shocks. When financial constraints are tightened—such as by increasing the cost of external financing or raising the idiosyncratic volatility of cash flows—it tends to amplify this sensitivity, resulting in a higher firm beta. This excess sensitivity is again captured by the marginal value of cash. An implication is that the standard textbook treatment for how to unlever a firm's beta is biased because it ignores the impact of cash holding changes on financial constraint.

It is worth emphasizing that the marginal value of cash is an intrinsically forward-looking quantity, which endogenously depends on a firm's current financial slack, its outside options, and the future uncertainties about investment opportunities and financing conditions. Only in two special cases can we directly pin down the marginal value of cash: when the firm is completely unconstrained, in which case the marginal value of cash equals one, or when the firm is on the margin of raising external funds, in which case the marginal value of cash is equal to the marginal cost of external financing. In all other cases, we need a dynamic framework.

Our review of the subject of corporate liquidity is complementary to the work by Almeida et al. (2014), Berg, Saunders & Steffen (2021), and Denis & Wang (2024). Our review focuses on the recent theoretical developments and emphasizes the dynamic perspective. We also examine the asset pricing implications of financial constraints. For a review on the broader topic of dynamic corporate finance, including liquidity management, investment, and financial policies, please see Strebulaev & Whited (2012).

In the remainder of this article, we begin by describing the static model and deriving its main predictions on the two margins in Section 2. We then proceed in Section 3 to the analysis of a simple dynamic model based on the q theory of investment (Hayashi 1982) but for a firm that faces external financing costs. We analyze how the firm optimizes along the two margins to determine its dynamic investment, corporate savings, and risk management policies. In Section 4, we discuss the asset pricing implications of corporate liquidity, and we end with discussions about new directions of research in Section 5. Section 6 concludes.

2. A STATIC MODEL

The static model we consider here is largely based on the work by Froot, Scharfstein & Stein (1993) and Kaplan & Zingales (1997). What we mainly add to their analyses is the key notion of the marginal value of cash (a marginal change in corporate savings) and how corporate investment is affected by any change in the availability of internal funds.

2.1. Model and Solution

Consider a firm that has an investment opportunity and is endowed with an initial stock of internal funds $w \geq 0$ at $t = 0$. We shall take it that the investment project is scalable but has decreasing returns to scale. We denote by I the size of the investment the firm undertakes. Let's begin with the simplest situation by assuming that this investment generates a deterministic cash flow $f(I) > 0$ at $t = 1$, which is increasing and concave in I: $f'(I) > 0$ and $f'(I) < 0$. The interest rate is normalized to zero.

2.1.1. Costly external financing. Consider the generic situation in which the firm's cash holding w is not large enough to finance its desired level of investment. It will then consider financing its desired investment with a combination of internal and external funds. However, raising external funds is costly. For brevity, we will not delve into the microfoundations of the source of these costs (asymmetric information, incentives, limited commitment) and simply postulate an external financing cost function, which takes the following form.

If the firm raises an amount e in external funds, it incurs a cost $C(e; \theta)$, where θ is a parameter that captures how costly external financing varies with market conditions. We shall take it that a higher parameter value θ corresponds to higher external financing costs. When $\theta = 0$, external financing is costless. We assume that the external financing cost function $C(e)$ is smooth, increasing, and convex: $C_e > 0$ and $C_{ee} \geq 0$, with $C(0) = 0$.

The firm's problem is to choose the level of its investment I and the amount of external financing e to maximize its profits:

$$\max_{I,e} \; f(I) + w - (I + C(e)), \tag{4}$$

where external financing e satisfies the following funding requirement:

$$e = \max\{I - w, 0\}. \tag{5}$$

The first two terms in Equation 4 are, respectively, the firm's revenues $f(I)$ and its cash holdings w. The third term is the total cost the firm incurs to generate revenues $f(I)$. This includes both the investment cost I and the cost of raising external financing $C(e)$. Equation 5 states that the amount of external financing e is nonnegative. If $I - w \leq 0$, there is no external financing, i.e., $e = 0$, and the firm is entirely internally financed. In that case, the firm can save the surplus funds internally from $t = 0$ to $t = 1$ at no cost and distribute them at $t = 1$. In any case, we are mainly interested in the situation where the firm cannot entirely cover its desired capital expenditures with w, i.e., when $I - w > 0$.

Let $P(w)$ denote the firm's value function for the optimization problem (Equations 4 and 5) and let $V(w) = P(w) - w$ denote the firm's enterprise value, which measures its ability to create value for shareholders. Under the Modigliani-Miller perfect capital market assumptions, the firm's ability to create value is independent of its financial structure, which here means that neither the internal-external financing margin nor the debt-equity margin are relevant. However, for a firm facing external financing costs, financial policies matter for value creation, and the enterprise value $V(w)$ captures value creation derived from both fundamentals (the production technology) and financing choices. In the dynamic formulation we lay out in the next section, we link the firm's enterprise value to Tobin's q.

2.1.2. First-best: Modigliani-Miller financing irrelevance.

When external financing is costless [$\theta = 0$ and, equivalently, $C(e; \theta) \equiv 0$], the first-best investment level is attained and given by the following first-order condition (FOC):[2]

$$f'(I^{FB}) = 1. \tag{6}$$

Importantly, in this first-best situation, I^{FB} maximizes the investment project's value and is independent of the firm's cash holding w. The firm's enterprise value, $V^{FB}(w)$, is given by $V^{FB}(w) = P^{FB}(w) - w = f(I^{FB}) - I^{FB}$, which is also independent of its cash holding w.

When external financing is costly but the firm has sufficient internal funds ($w \geq I^{FB}$), the firm can still invest at the first-best level, $I = I^{FB}$. This is simply because the firm is financially unconstrained and does not need to rely on any external financing sources.

2.1.3. Investment and valuation for a financially constrained firm.

Next, we show how costly external financing affects a financially constrained firm's investment choice I and its enterprise value $V(w)$. In this static model, a firm is financially constrained if its internal funds w are insufficient to finance I^{FB}, i.e., when $w < I^{FB}$. The optimal investment as a function w, $I^*(w)$, is then determined by the following FOC for the case where external financing is positive:

$$f'(I^*(w)) = 1 + C'\left(e^*(w)\right) = 1 + C'\left(I^*(w) - w\right). \tag{7}$$

[2]We can solve the firm's first-best problem in two steps. First, we choose I to solve $\max_I f(I) + w - I$, and then we obtain e by using Equation 5. The second-order condition is satisfied as $f''(I) < 0$.

The firm first exhausts its internal funds and then taps the costly external funding by raising $e^*(w) = I^*(w) - w \geq 0$.[3] Because the investment project has decreasing returns to scale $[f''(I) < 0]$, and because external financing is costly $[C'(e) > 0]$, the investment choice $I^*(w)$ is strictly less than the first-best level $I^*(w) < I^{FB}$ when the firm is financially constrained.

2.1.4. Marginal value of cash $P^*_w(w)$. Substituting for $I^*(w) = e^*(w) + w$ into the objective function (Equation 4), we obtain

$$P^*(w) = f(e^*(w) + w) - [e^*(w) + C(e^*(w))]. \quad 8.$$

Applying the envelope theorem to Equation 8 and using the FOC (Equation 7), we obtain the following result for the marginal value of cash:

$$P^*_w(w) = f'(I^*(w)) = 1 + C'(e^*(w) \geq 1. \quad 9.$$

Like Equation 7, Equation 9 equates the marginal value of investment, $f'(I^*(w))$, with the marginal cost. The key difference from first-best is that the marginal cost of investing another dollar is now determined by the marginal value of cash, which is above one under financial constraint. When the firm lacks sufficient internal funds, it is underinvesting in the positive net present value (NPV) project, and any additional dollar in internal funds reduces the amount by which the firm underinvests. Stated differently, any additional dollar in corporate savings allows the firm to create value at the margin, which is captured by the positive net shadow value of cash: $V^*_w(w) = P^*_w(w) - 1 > 0$.

Given the assumption that the external financing cost is strictly increasing in funds raised, i.e., $C_e > 0$, the marginal value of cash $P^*_w(w)$ is above one whenever the firm is financially constrained, i.e., $w < I^{FB}$. Furthermore, if the financing cost is convex, i.e., $C_{ee} > 0$, the marginal value of cash will be higher the lower the firm's cash holding w (and the larger the external financing needs), which further exacerbates the underinvestment problem (as implied by Equation 9).[4] This link between the marginal value of cash and investment is one of the most robust implications of the presence of external financing costs. It highlights the marginal value of cash as a key statistic for the wedge between a firm's external versus internal funding margin and helps explain how financing constraints and cash holdings affect corporate investment.

2.2. Model Predictions

As pared down as this simple static model is, it yields a rich set of predictions on how the enterprise value, the marginal value of cash, corporate investment, and the cash sensitivity of investment vary with corporate savings w. We illustrate these predictions in **Figure 4**, which plots, respectively, the enterprise value $V^*(w)$, the net marginal value of cash $V^*_w(w)$, optimal investment $I^*(w)$, and the cash sensitivity of investment $I^*_w(w)$ as a function of w. The solution plotted in **Figure 4** is obtained for the following parameter choices: We assume that the production function $f(I)$ takes the simple functional form $f(I) = I^\alpha$ and that the external financing cost function takes the form $C(e; \theta) = \theta_1 \cdot e + \theta_2 \cdot e^2/2$. We have set the following parameter values for these two functional forms: $\alpha = 0.7$ for the slope and curvature of the production function, and $\theta_1 = 0.09$, $\theta_2 = 0.04$ for the external financing cost function.

For these parameter values, the first-best investment level is given by $I^{FB} = \alpha^{\frac{1}{1-\alpha}} = 0.30$, and the first-best enterprise value is $V^{FB}(w) = (1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}} = 0.13$. They are independent of w, as depicted in **Figure 4** with the red dashed horizontal lines. If the firm is financially constrained

[3] It is possible that external financing is zero for a financially constrained firm. This happens when the marginal product $f'(I)$ is lower than the marginal cost of investing purely with internal funds: $1 + C'(0)$.

[4] As we will explain below, in dynamic settings, convexity of external financing costs is neither sufficient nor necessary for the marginal value of cash to be decreasing in cash holding.

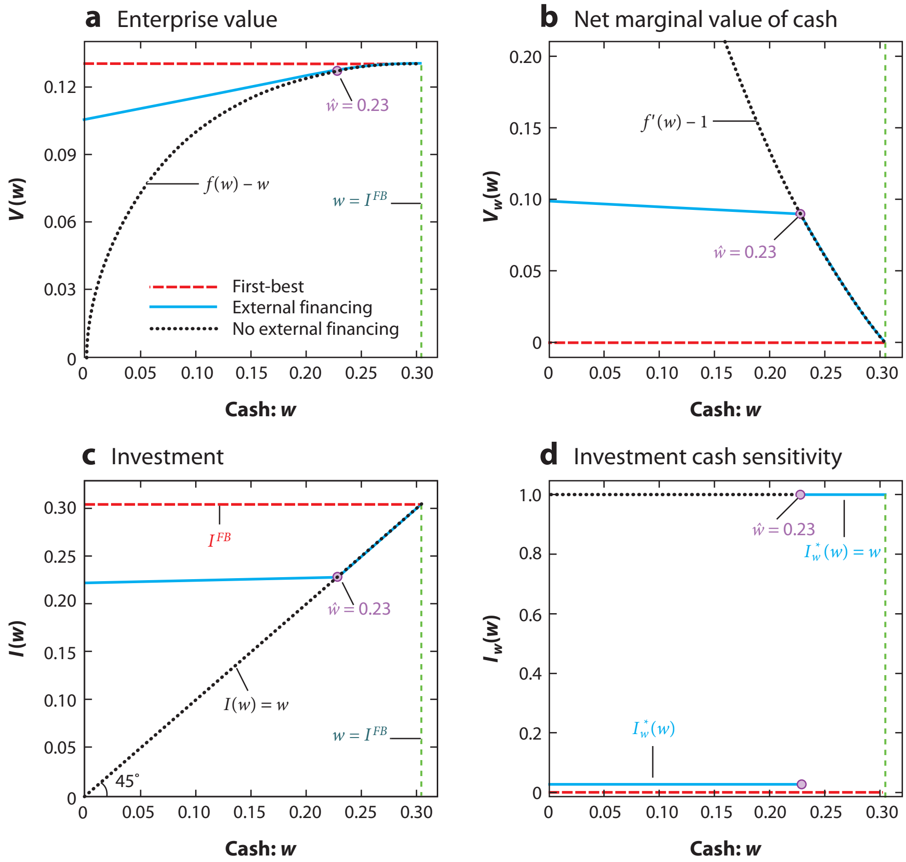

Figure 4

Investment and value for financially constrained firms: a static model. This figure shows the (*a*) enterprise value, $V(w)$; (*b*) net marginal value of cash, $V_w(w)$; (*c*) investment, $I(w)$; and (*d*) investment cash sensitivity, $I_w(w)$, as functions of cash holding w. The panels show the value of each variable under three cases: the first-best case (*red dashed line*), the case with external financing (*blue solid line*), and the case without external financing (*black dotted line*). The green dashed vertical lines mark the boundary where a firm becomes unconstrained, while purple circles mark the threshold for cash above which no external financing is used.

$(w < I^{FB})$, it sets its investment policy $I^*(w)$ as follows: When cash holdings are very low so that $w \in (0, \widehat{w})$, where $\widehat{w} = \left(\frac{\alpha}{1+\theta_1}\right)^{\frac{1}{1-\alpha}} = 0.23$, the investment FOC,

$$\alpha(I^*(w))^{\alpha-1} = 1 + \theta_1 + \theta_2(I^*(w) - w),$$

pins down $I^*(w)$, and the amount of external financing is given by $e^*(w) = I^*(w) - w$. For higher levels of cash holdings such that $w \in (\widehat{w}, I^{FB}) = (0.23, 0.30)$, the firm finances its investment by using internal funds only and sets $I^*(w) = w$. This corner solution obtains because the net marginal value of investment is less than the proportional cost of raising external funds $\theta_1 = 0.09$ (see the blue solid lines in **Figure 4**).

Notice that $V^*(w)$, which is represented by the blue solid line in **Figure 4*a***, is increasing and concave. In the region $w \in (0, \widehat{w})$, it is optimal to raise external financing. This is why the blue solid line in **Figure 4*a*** is above the black dotted line, which represents the enterprise value $V(w)$ for a firm without access to external financing. The wedge between the solid blue and dotted black lines is the value of external financing, which is zero in the region $w \in (\widehat{w}, I^{FB}) = (0.23, 0.30)$. As **Figure 4*b*** illustrates, the net marginal value of cash decreases with w, and as **Figure 4*c*** shows, firm investment increases with w.[5] Finally, the investment cash sensitivity $I_w^*(w)$ is strictly positive but significantly less than one in the region $w \in (0, \widehat{w}) = (0, 0.23)$ and equals one in the region $w \in (\widehat{w}, I^{FB}) = (0.23, 0.30)$. And if $w > I^{FB}$, it obviously drops to zero, so that $I_w^*(w)$ is both nonlinear and nonmonotonic with respect to cash holdings w.[6]

2.3. Limitations of the Static Model

The static model we have just outlined provides some key insights on how costly external financing affects corporate investment. However, there are fundamental limitations to the static setting. First, by assumption, the firm is liquidated at the end of the period and all proceeds are paid out. Thus, the model cannot provide meaningful predictions regarding the payout policy, which depends on the intertemporal trade-off between immediate payout and saving cash for the future.

Second, the model is limited in providing guidance on the timing or amount of external financing. In this model, the timing of external financing always coincides with that of investment (at time 0), while the amount of funds raised is simply equal to the gap between investment and internal funds. In reality, both decisions are dynamic and state-contingent, and they will no longer be mechanically tied to investment when we allow for active choices on cash savings.

Third, the static model does not fully capture the variation in the (net) marginal value of cash, which reflects the degree of financial constraints. From Equation 9, we can see that the marginal value of cash in this model is either zero, if the firm does not raise any external funds at $t = 0$, or it is directly pinned down by the marginal cost of external financing $C'(e^*)$. In reality, significant fund raising (through either debt or equity) happens rather infrequently for most firms. However, this does not mean these firms are unconstrained. To determine the marginal value of cash in such states would require us to take into account the current financial slack, the uncertainty about future liquidity needs, and the uncertainty about future financing opportunities. We address these questions in a dynamic framework in the next section.

3. A DYNAMIC FRAMEWORK

We take the approach of analyzing corporate investment dynamics under financial constraints by introducing financing frictions inside the classic q theory of investment. In this parsimonious framework, the marginal product of capital (known as marginal q) and the marginal value of cash are the two key determinants of corporate investment and savings dynamics. Below, we describe a simplified version of such a model, which is based on Bolton, Chen & Wang (2011).

Starting from the q theory of investment, the firm's operations are described by its physical capital stock, K_t, which evolves as follows:

$$\mathrm{d}K_t = (I_t - \delta K_t)\mathrm{d}t, \qquad 10.$$

where I_t is investment and δ is the depreciation rate of the capital stock. Total capital expenditure costs equal the sum of investment outlays and adjustment costs: $I + G(I, K)$, where the adjustment cost is increasing and convex in I: $G_I(I, K) > 0$ and $G_{II}(I, K) > 0$.

[5] Note that $V^*(w)$ is continuously differentiable, including when $w = \widehat{w}$, but investment $I^*(w)$ is continuous and not differentiable at $w = \widehat{w}$.
[6] The investment cash sensitivity is $I_w^*(w) = \theta_2 / \left(\theta_2 - \alpha(\alpha - 1)(I^*(w))^{\alpha-2}\right) > 0$ and $V_w^*(w) = \theta_1 + \theta_2(I^*(w) - w)$ is nonlinear in w in the $w < \widehat{w}$ region. In the internal financing region where $w \in (\widehat{w}, I^{FB})$, $I^*(w) = w$ and $V^*(w) = f(w) - w$, implying $I_w^*(w) = 1$ and $V_w^*(w) = \alpha w^{\alpha-1} - 1$, respectively.

At any time t, shareholders can decide whether to continue the operation or to close down the firm. If the firm is liquidated, the scrap value of the assets is $\ell K_t + W_t$, where $\ell > 0$ is the liquidation value per unit of capital stock and W_t is the stock of cash the firm has accumulated up to time t. If shareholders decide to continue operations, the firm produces a (risk-adjusted) operating revenue of $K_t \mathrm{d}A_t$, where $\mathrm{d}A_t$ is the firm's (risk-adjusted) productivity over time interval $\mathrm{d}t$.

The simplest form of cash flow uncertainty to consider is to assume that productivity shocks are governed by the following law of motion:

$$\mathrm{d}A_t = \mu \mathrm{d}t + \sigma \mathrm{d}\mathcal{Z}_t, \qquad 11.$$

where $\mathcal{Z}_t$ is a standard Brownian motion (after a suitable risk adjustment under the risk-neutral measure).[7] The (risk-adjusted) expected productivity is equal to μ, and the volatility is σ per unit of time. Under this law of motion, productivity shocks $\mathrm{d}A_t$ are effectively identically and independently distributed (i.i.d.) over time.

The (risk-adjusted) free cash flow generated by the firm's operations over the interval of time $\mathrm{d}t$ is given by

$$\mathrm{d}Y_t = K_t \mathrm{d}A_t - (I_t + G(I_t, K_t))\mathrm{d}t. \qquad 12.$$

Under the pure q theory of investment (without any external financing costs), shareholders' objective is to maximize shareholder value by choosing investment outlays $\{I_s;\ s \in (t, \tau)\}$, and a liquidation stopping time τ to maximize the expected discounted value of free cash flows:

$$\mathbb{E}_t \left[\int_t^{\tau} e^{-r(s-t)} \mathrm{d}Y_s + e^{-r(\tau - t)} \left(\ell K_\tau + W_\tau \right) \right]. \qquad 13.$$

Note that since the expectation $\mathbb{E}_t[\,\cdot\,]$ is taken under the risk-neutral measure, any risk premium shareholders might demand has been taken into account. We denote by $P(K_t, W_t)$ the shareholder value associated with capital stock K_t and corporate savings W_t.

Shareholders generally prefer a bird in the hand than in the bush, so to speak. This basic preference for cash paid out is reflected in the model through the introduction of a small cash-carrying cost $\lambda > 0$ for retained earnings. Under this assumption, the firm's cash management equation is given by:

$$\mathrm{d}W_t = \underbrace{K_t \mathrm{d}A_t}_{\text{revenue}} - \underbrace{(I_t + G(I_t, K_t))\,\mathrm{d}t}_{\text{capital expenditure}} + \underbrace{(r - \lambda)\, W_t \mathrm{d}t}_{\text{net interest income}} + \underbrace{\mathrm{d}H_t - \mathrm{d}U_t}_{\text{net issuance}}. \qquad 14.$$

The first term on the right-hand side of Equation 14 represents the firm's realized revenue $K_t \mathrm{d}A_t$. The second term, $(I_t + G(I_t, K_t))\mathrm{d}t$, represents the firm's capital expenditures. The third term, $(r - \lambda)W_t \mathrm{d}t$, represents the net interest income on the firm's cash holdings W_t. The fourth term, $\mathrm{d}H_t \geq 0$, denotes any net external financing raised by the firm, and the fifth term, $\mathrm{d}U_t \geq 0$, denotes any cash payout to shareholders (both $\mathrm{d}H_t$ and $\mathrm{d}U_t$ are nonnegative at all times). The cash-carrying cost λ may reflect either a smaller investment opportunity set for the firm than is available to shareholders or an agency cost. When raising external financing is costly, it is generally desirable for the firm to retain earnings. But, there is a limit to how much the firm may want to save when it incurs a positive cash-carrying cost.

Shareholder value $P(K_t, W_t)$ depends on at least two state variables, capital stock K_t and cash holding W_t. This means that in general, the firm's dynamic investment and financing problem is (at least) two-dimensional. Dynamic optimization problems in two dimensions can be tractable,

[7]Mathematically, the risk adjustment is via the market price of risk in the stochastic discount factor, which connects the physical measure to the risk-neutral measure.

and a number of such models exist in the literature.[8] Still, there is a higher barrier to entry for researchers as well as a sacrifice of the transparency of underlying economics. This is why it can be valuable to adopt the necessary, albeit restrictive, assumptions to guarantee the homogeneity of the value function, which allows one to reduce the model to a one-dimensional problem.

Concretely, we impose a constant-return-to-scale assumption for the production function, as shown in Equation 12, which is also referred to as the A-K technology. This assumption, together with the homogeneity of the external financing costs and investment adjustment costs, ensures that the value function $P(K_t, W_t)$ is homogeneous of degree one. Thus, we can define the relevant variables and functions in per-unit-of-capital terms, including the cash-capital ratio $w_t = W_t/K_t$, investment-capital ratio $i_t = I_t/K_t$, adjustment cost-capital ratio $g(i_t) = G(I_t, K_t)/K_t$, and firm value-capital ratio $p(w_t) = P(K_t, W_t)/K_t$. The firm's average q (or market-to-book ratio) is defined as the firm's enterprise value, $P(K_t, W_t) - W_t$, normalized against the capital stock K_t,

$$q_a(w_t) = \frac{P(K_t, W_t) - W_t}{K_t} = p(w_t) - w_t. \tag{15}$$

In a Modigliani-Miller world with perfect capital markets, there is no meaningful distinction between internal and external funds, and the firm's average q is independent of its financing policies. We denote average q in this situation as the first-best q:

$$q_a(w) = q^{FB} = \max_i \frac{\mu - (i + g(i))}{r + \delta - i}. \tag{16}$$

The twin assumptions of i.i.d. productivity shocks and constant returns to scale (homogeneity of degree one) imply that the first-best q is a constant. Below, we start with the simplest version of Bolton, Chen & Wang (2011): the self-financing model.

3.1. Self-Financing Model

We start by considering the setting with the most extreme form of financial constraint, in which the firm does not have access to any external funding and is entirely self-sufficient. In this situation, the firm can only service its liabilities out of internally generated cash flows and corporate savings. Each dollar spent on investment entails an opportunity cost: There is one dollar less available in the future to cover any operating costs or finance any new investment opportunities. This opportunity cost is greater the smaller the accumulated cash holdings of the firm. When the firm runs out of cash it must be liquidated. At the other extreme, when the firm is flush with cash it may want to dispose of some of that cash by paying it out to shareholders.

Thus, depending on the size of corporate savings W_t, the firm's investment and financial policy can be sorted into three broad regions: a liquidation region (when $W_t = 0$), an internal financing region, and a payout region. When the operating asset generates strictly positive expected earnings, as we assume, the going-concern value is strictly positive, so that it is not in shareholders' interest to liquidate the asset unless they have to (should the firm run out of cash and be unable to cover any realized losses).

Since the firm incurs a positive cost of carrying cash ($\lambda > 0$), corporate savings can reach a point when it is better to pay out any additional earnings rather than to retain them. We denote this point by the (endogenous) payout boundary $\overline{W}(K)$. With no additional costs, payout is an optimal instantaneous control problem: It will be made in infinitesimal amounts, as the cash-holding process (Equation 14) is reflected at the boundary $\overline{W}(K)$.

[8]Such an approach is much more common in the discrete time setting. Examples of continuous-time models include Anderson & Carverhill (2012), Bolton, Wang & Yang (2019a), and Kakhbod et al. (2023).

Consider the shareholders receiving payout amount $\overline{W}(K) - W$ right after reaching the payout boundary, which brings the cash holding back to $W \leq \overline{W}(K)$. The shareholder value is given by

$$P(K, W) = P(K, \overline{W}(K)) - (\overline{W}(K) - W).$$

Taking the limit as $W \to \overline{W}(K)$, the value-matching condition above becomes

$$P_W(K, \overline{W}(K)) = 1. \qquad 17.$$

This condition is intuitive: It states that the marginal value of cash equals one at the payout boundary. Notice that no optimization is involved in deriving this condition.

Next, the optimal payout boundary $\overline{W}(K)$ is pinned down by a super contact condition (Dumas 1991),

$$P_{WW}(K, \overline{W}(K)) = 0. \qquad 18.$$

Intuitively, the optimality for the payout boundary $\overline{W}(K)$ requires that the firm be indifferent between holding onto an extra dollar of cash inside the firm and paying the dollar out to shareholders. This means that the marginal value of cash should be equal to one both at the time of payout and the instant afterward.[9]

In the internal financing region, where $0 < W_t < \overline{W}(K)$, the firm retains all its earnings since a dollar saved inside the firm is worth more than a dollar in shareholders' pockets. In this region, therefore, the firm optimally sets $dU_t = 0$. Moreover, its optimal level of investment I is the solution to the following recursive maximization problem defined by the Hamilton-Jacobi-Bellman equation for firm value, $P(K, W)$:

$$rP(K, W) = \max_I \; (I - \delta K) P_K + [(r - \lambda)W + \mu K - I - G(I, K)]P_W + \frac{\sigma^2 K^2}{2} P_{WW}. \qquad 19.$$

Equation 19 is a key equation of the dynamic model. It has a natural interpretation. The first term on the right-hand side represents the marginal effect of additional capital $(I - \delta K)$ on firm value $P(K, W)$, where P_K is the marginal value of capital (known as marginal q in the q theory of investment). The second term captures the marginal effect of additional corporate savings on firm value $P(K, W)$:

$$\underbrace{(r - \lambda)W}_{\text{net interest income}} + \underbrace{\mu K}_{\text{expected cash flow from operations}} - \underbrace{(I + G(I, K))}_{\text{capital expenditure}},$$

where P_W is the marginal value of cash. The last term reflects the effect of uncertainty. Since free cash flows are random, the accumulation of corporate savings is uncertain. The curvature of the value function $P(K, W)$, as measured by P_{WW}, captures the effect of this uncertainty on firm value.

There is a vast literature on investment in macroeconomics and finance. Most of this literature has focused on marginal q, as captured by $P_K(K, W)$ in this model. This is because marginal q is often taken to be a measure of investment opportunities, which is expected to determine investment. However, an exclusive focus on marginal q is misleading, because it abstracts from any effects on investment through changing financing conditions firms are exposed to. The simple model described here adds this second dimension and highlights how both marginal q and the marginal value of cash P_W are key drivers of investment.

[9]For a heuristic derivation of the super contact condition, the total derivative of both sides of Equation 17, with respect to $\overline{W}(K)$, results in $P_{WW}(K, \overline{W}(K)) + \frac{\partial}{\partial \overline{W}(K)} P_W(K, \overline{W}(K)) = 0$. The second term is equal to zero due to the optimality of $\overline{W}(K)$, which then leads to Equation 18.

The FOC for investment implied by Equation 19 is:[10]

$$1 + G_I(I, K) = \frac{P_K(K, W)}{P_W(K, W)}. \tag{20}$$

Investment is not driven by marginal q but by the ratio of marginal q to the marginal value of cash. When it increases capital by one unit, the firm incurs a capital expenditure cost in dollars of $1 + G_I(I, K)$. Crucially, to pay for $1 + G_I(I, K)$, the firm must dip into its cash savings, which have a marginal opportunity cost $P_W(K, W)$. The marginal value of cash exceeds one for all $0 \leq W_t < \overline{W}(K)$ and is higher the lower is W_t. In effect, the firm's marginal cost of investing is therefore equal to $(1 + G_I(I, K))P_W(K, W)$, which at the optimal level of investment equals the marginal benefit of investment $P_K(K, W)$. We highlight this key first-order optimality condition in words as follows:

marginal q = marginal (physical) cost of investing × marginal value of cash.

Note that, in the special case where the firm is not financially constrained (or faces no external financing costs), $P_W = 1$. The optimality condition for investment (Equation 20) then boils down to the standard condition in the q theory of investment,

$$1 + G_I(I, K) = P_K(K, W).$$

That is, with convexity capital adjustment costs, the marginal (physical) cost of investing is equal to the marginal q.

We can express the key Equations 19 and 20 in terms of the variables $w = W/K$ and $i = I/K$ by exploiting the homogeneity of degree one of the model in terms of the capital stock K. Recall that we have defined the functions $g(i) = G(I, K)/K$ and $p(w) = P(K, W)/K$, so that the optimal investment-capital ratio, $i(w)$, is pinned down by the ratio of marginal q, $q_m(w)$, and the marginal value of cash, $p'(w)$:

$$1 + g'(i(w)) = \frac{q_m(w)}{p'(w)} = \frac{p(w)}{p'(w)} - w, \tag{21}$$

since $q_m(w) = p(w) - p'(w)w$. Replacing $P(K, W) = p(w)K$ in Equation 19, we can immediately verify that the value $p(w)$ of a unit of capital is then given by

$$rp(w) = (i(w) - \delta)\, p(w) + \mu_w(w)p'(w) + \frac{\sigma^2}{2}p''(w), \tag{22}$$

where $i(w)$ is given in Equation 21, $\mathrm{d}w_t = \mu_w(w_t)\mathrm{d}t + \sigma\,\mathrm{d}Z_t$, and $\mu_w(w)$ is the drift of w given by

$$\mu_w(w) = (r - \lambda)w + \mu - i(w) - g(i(w)) - w\,(i(w) - \delta). \tag{23}$$

At the optimal payout boundary $\overline{w}$, the marginal value of cash is equal to one (see Equation 17) and the sensitivity of $p'(w)$ is zero (as implied by Equation 18):

$$p'(\overline{w}) = 1 \quad \text{and} \quad p''(\overline{w}) = 0. \tag{24}$$

Finally, $p(0) = \ell$, as the firm is liquidated when it runs out of cash.

Many intuitive insights can be gained from a depiction of how $p(w)$ and $i(w)$ vary with the ratio of corporate savings to capital, w.[11] In **Figure 5*a***, we show that $p(w)$ is increasing and concave in

[10]Note that the second-order condition is automatically satisfied as $G_{II}(I, K) > 0$.

[11]We use a quadratic adjustment cost function: $g(i) = \theta i^2/2$ and the same parameter values as in Bolton, Chen & Wang (2011) for our numerical solution.

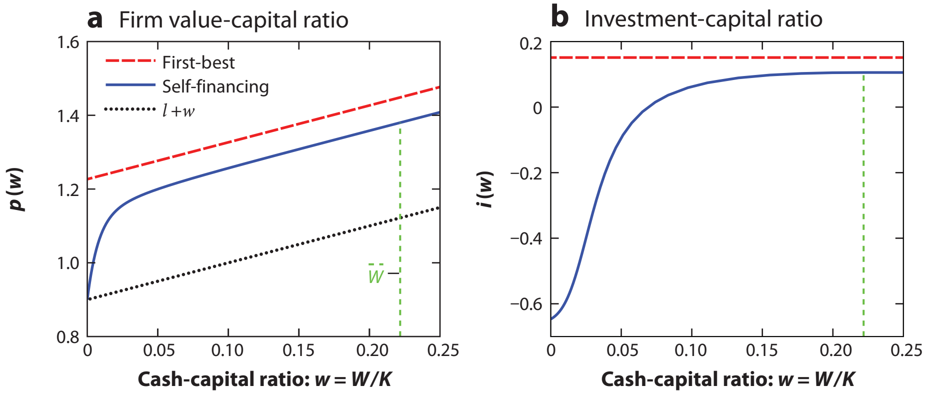

Figure 5

A dynamic self-financing model. Panel *a* shows the firm value-capital ratio, $p(w)$, under the first-best case (*red dashed line*) and the self-financing case (*blue solid line*). The black dotted line, $l + w$, denotes the liquidation value. Panel *b* does the same for the investment-capital ratio, $i(w)$. The green dashed vertical lines mark the payout boundaries. Figure adapted from Bolton, Chen & Wang (2011).

w. Note that the liquidation value $p(0) = \ell = 0.9$ pulls down $p(w)$, which generates a very large curvature near $w = 0$. This means that the marginal value of cash is high when w is close to zero and that the firm is very averse to risk with respect to changes in w, as even a small reduction in w could mean death for the firm. The marginal value of cash $p'(w)$ exceeds one when a firm is financially constrained, and $p'(w)$ increases as w decreases. In **Figure 5*b***, we show that low corporate savings causes the firm to underinvest: $i(w) < i^{FB} = 15.1\%$. Moreover, underinvestment is very severe when w approaches 0. To be sure, far from investing, the firm actually divests some of its operating assets in an effort to generate cash. The firm disposes of more than 60% of its capital stock at fire-sale prices when it is about to run out of cash. Asset sales are then the only option left for the firm to raise cash and increase the likelihood of survival.

This model generates a stochastic investment process, given that $i(w_t)$ varies with w_t, which follows a diffusion process with drift given in Equation 23 and volatility σ. Importantly, this stochastic process for investment is not driven by shocks to the firm's investment opportunities, a common explanation in the literature. Rather, it is the financial performance of the firm, which affects its corporate savings, that causes the firm to scale back or increase its capital expenditures. Another noteworthy prediction is that the cash sensitivity of investment $i'(w)$ is nonmonotonic in w (a result also derived by Kaplan & Zingales 1997 in a static model). When money is tight (w is close to zero), a small increase in w allows the firm to reduce the scale of its assets sales, and when money is more abundant (w is closer to the payout boundary $\overline{w}$), a small increase in w slows down the rate of increase in investment. All these rich time series implications are absent in static models.

3.2. Costly External Financing, Credit Line, and Hedging

We now generalize the self-financing model of Section 3.1 by adding the possibility of external financing through equity issuance, borrowing through a line of credit, and also hedging positions. These decisions interact with the financial constraint in rich ways, which have important practical implications.

3.2.1. Equity issuance. We now assume that the firm can raise money from equity markets by incurring a fixed and a proportional cost of equity issuance. We preserve the homogeneity property of the model by assuming that the quasi-fixed equity-issuance cost Φ is proportional to the firm's size as measured by its installed capital, $\Phi = \phi K$, with $\phi > 0$. This means that the firm will not grow out of its fixed issuance costs. We denote the marginal cost of issuing equity as $\gamma > 0$. Multiple empirical studies of corporate equity issuance have found that both fixed and variable underwriting and other related issuance costs are substantial. Because of these costs, corporations tend to avoid equity issuance if they can. When they do tap equity markets, corporations do so on an intermittent and lumpy basis as a way of minimizing overall issuance costs.

Since equity issuance is costly and the costs of issuance are constant over time, it is optimal for the firm to delay issuance as much as it can, i.e., tap equity markets only when it runs out of cash. At that point, it has two options, liquidation or recapitalization. If equity issuance costs are not too high, as we assume here, it is preferable for the firm to recapitalize rather than fold.

In the presence of fixed costs, equity issuance is an optimal impulse control problem: Conditional on issuing equity, the firm will want to raise a discrete amount. Denoting by $M_t = mK_t$ the net amount raised through an equity issue, we characterize equity issuance at the point when the firm runs out of cash with the following two conditions:

$$p(0) = p(m) - \phi - (1+\gamma)m \quad \text{and} \quad p'(m) = 1 + \gamma. \qquad 25.$$

These two conditions now replace the boundary condition $p(0) = \ell$ in the self-financing model.[12] We refer to m as the target amount of net funding through the equity issue. The first condition in Equation 25 states that firm value [netting out total equity issuance costs $\phi + (1+\gamma)m$] is continuous before and after the equity issue. The second condition in Equation 25 states that the last dollar raised through the equity issue has a marginal value of cash $p'(w)$ that is equal to the marginal cost of issuance $1 + \gamma$.

The firm's dynamic investment, saving, and equity issuance policy is then characterized by Equations 21, 22, and 24 for investment $i(w)$, firm value $p(w)$, and payout policies, respectively, and by the boundary conditions for the firm's equity issuance timing and quantity decisions given by Equation 25. As long as the obtained value of $p(0)$ is larger than ℓ, it is optimal for the firm to recapitalize by tapping external equity markets rather than liquidating.

Many intuitive insights can again be gained from a depiction of how $p(w)$ and $i(w)$ vary with the ratio of corporate savings to capital, w. Moreover, the comparison of **Figure 5** with **Figure 6** yields additional insights. In **Figure 6*a***, we show that $p'(w)$ is positive and decreasing in w, which implies that $p(w)$ is increasing and concave in w. We display two solutions, one when the fixed issuance cost is assumed to be zero ($\phi = 0$) and the other when it is given by $\phi = 1\%$. For both solutions, we assume that the marginal cost of issuance is $\gamma = 6\%$.

Obviously, firm value when $\phi = 0$ (the black dotted line) is higher than when $\phi = 1\%$ (the blue solid line). Interestingly, the marginal value of cash stays close to 1 when $\phi = 0$. As a result, the value function is almost linear and close to the first-best solution (the red dotted line in **Figure 5*a***). It is not quite as high because every so often the firm must tap equity markets and incur the marginal issuance cost γ.

When the firm faces fixed issuance costs ($\phi = 1\%$), the marginal value of cash can become substantially higher than 1 when the cash holding is low. In this case, the firm raises a lumpy amount of funds m when it taps equity markets, which is indicated by the green vertical line on the left side of **Figure 6*b***. In contrast, when it faces no fixed costs of issuance ($\phi = 0$, so that it only pays the marginal cost γ), it only raises the amount it needs to cover any losses.

Another striking difference between the two solutions is that, when $\phi = 1\%$, the firm seeks to accumulate more savings by delaying payout. This can be seen by comparing the payout boundary

[12]When there is no fixed issuance cost ($\phi = 0$), equity issuance is small and only occurs when $w = 0$.

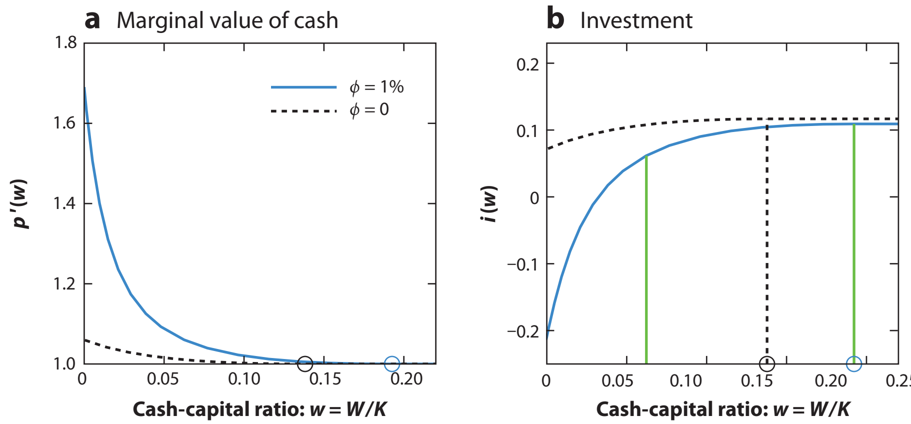

Figure 6

An external equity financing model. Panel *a* shows the marginal value of cash, $p'(w)$, under the case with fixed issuance cost ($\phi = 1\%$) and without ($\phi = 0$). Panel *b* does the same for the investment-capital ratio, $i(w)$. The green vertical lines and the two circles on the x-axis denote the payout boundaries under these two cases ($\phi = 1\%$, *blue circles*; $\phi = 0$, *black circles*). Figure adapted from Bolton, Chen & Wang (2011).

when $\phi = 0$ (the black dotted vertical line) to the payout boundary when $\phi = 1\%$ (the green vertical line on the right side of **Figure 6*b***).

We also see in **Figure 6*b*** that investment $i(w)$ when $\phi = 0$ is almost invariant to changes in w. It is not quite close to the first-best investment level, as can be seen by comparing the black dotted plot in **Figure 6*b*** to the red dotted line in **Figure 5*b***, because the firm every so often must raise costly equity. In contrast, when $\phi = 1\%$, we see that $i(w)$ is very sensitive to changes in w. Compared with the self-financing model, being able to issue equity (even when $\phi = 1\%$) increases firm value $p(w)$, lowers the marginal value of cash $p'(w)$, and mitigates underinvestment.

3.2.2. Related work. Riddick & Whited (2009) study a dynamic model of cash holdings and investment in discrete time. They characterize the marginal value of cash (referred to as the shadow value of cash balance) (see their equation 6) under two situations. First, in the payout region, the firm is completely unconstrained. In this case, the marginal value of cash equals one (subject to the correction on the effect of taxes). Second, when the firm is on the margin of raising external funds, the marginal value of cash is pinned down by the marginal cost of external financing (they model the costs of external equity financing as linear-quadratic). At all other times, which they refer to as "the region of financial inertia" (Riddick & Whited 2009, p. 1736) it is not as straightforward to characterize the marginal value of cash or study its implications on firm decisions due to the lack of tractability.

In the continuous-time setting, Decamps et al. (2011) also develop a firm liquidity and risk management problem. Their model does not feature corporate investment and hence is silent on the interactions between investment and liquidity management. Gryglewicz (2011) presents a tractable model of a firm that optimally chooses capital structure, cash holdings, dividends, and default while facing cash flows with both short-term liquidity shocks and uncertainty about long-term profitability. This model also abstracts away from investment and treats cash flows as exogenous. Anderson & Carverhill (2012) study a continuous-time model of corporate liquidity in the presence of long-term debt. They allow profitability shocks to be non-i.i.d., and they consider investment in the form of a real option. Due to the lack of tractability, they also do not characterize the marginal value of cash in the model.

3.2.3. Credit line. Our focus so far has been on the margin that is commonly suppressed in textbooks, the choice between internal and external financing and the optimal asset mix between liquid (cash) and illiquid (operating) assets. We now introduce another margin, the mix between debt and equity. We do this in a minimal way by adding another external financing option, a credit line commitment. This is an important source of external financing for firms in practice (see, e.g., Sufi 2009), in particular for small businesses, and its importance has risen even more since the global financial crisis (GFC) of 2008–2009 (see Berg, Saunders & Steffen 2021), including during the COVID-19 pandemic (see, e.g., Acharya & Steffen 2020; Greenwald, Krainer & Paul 2020).

We introduce the credit line commitment into the external financing model of Section 3.2.1 as follows. Let $\alpha > 0$ denote the spread between the interest rate charged by the bank on any amount drawn down from the line of credit and the risk-free rate r, and let the credit line limit be cK_t, where $1 > c > 0$ is an overcollateralization parameter that we set exogenously.

Since the credit line involves a positive external financing cost α, the firm will only choose to borrow once it has exhausted all internal funds. There is a pecking order of financing in this dynamic model as in Myers & Majluf (1984): The firm uses internal funds first, then relies on external debt financing, and as a last resort, relies on external equity financing. With the addition of the credit line facility, a new region for the state variable w is added, the credit region where $0 > w > -c$. In this credit region, $p(w)$ satisfies the following ordinary differential equation (ODE):

$$rp(w) = (i(w) - \delta)(p(w) - wp'(w)) + \left[(r+\alpha)w + \mu - i(w) - g(i(w))\right] p'(w) + \frac{\sigma^2}{2} p''(w). \qquad 26.$$

Comparing Equations 22 and 23 to Equation 26, we observe that the only difference between the two ODEs is the addition of the spread α in Equation 26. Access to a line of credit gives the firm more financial slack. It allows the firm to postpone a costly equity issue when it has run out of internal funds. It is only when the line of credit has been fully drawn down that the firm recapitalizes by issuing equity. This recapitalization allows the firm to delever and to put itself on a stronger financial footing to be able to rebuild its depleted corporate savings.

The following equations characterize the boundary conditions at $w = -c$ when the firm has exhausted its credit line:

$$p(-c) = p(m) - \phi - (1+\gamma)(m+c) \quad \text{and} \quad p'(m) = 1 + \gamma. \qquad 27.$$

Comparing Equations 25 and 27, we observe that the only difference between these boundary conditions is the addition of the credit line limit c. Given that the line of credit gives the firm more financial slack, it can afford to pay out accumulated savings sooner, so that the payout boundary $\overline{w}$ is lower for a firm with access to a credit line. Also, the marginal value of cash $p'(w)$ is lower and investment $i(w)$ is higher at all levels of w.

There is a surprising and striking prediction from this model concerning the relation between marginal q and investment. The standard prediction in the literature is that a higher marginal q translates into higher investment. The simple intuition behind this prediction is that higher marginal q means better investment opportunities, which invite more investment. However, as **Figure 7** shows, marginal q decreases with w in the credit region (i.e., when $w < 0$), while investment increases with w at all times. That is, marginal q and investment $i(w)$ move in opposite directions in the credit region. In contrast, average q is still monotonically increasing with w.

The economic logic behind this seemingly bizarre prediction is that, in the presence of financial constraints, marginal q not only conveys information about investment opportunities but also reflects the impact of changes in capital stock on the credit constraints. In the credit region, an additional unit of capital has the effect of relaxing the credit line limit (as an increase in K moves $w = W/K$ away from the credit limit $-c$ and closer to zero), and this effect is particularly strong when w is close to the credit limit. This is why marginal q is highest when $w = -c$ and decreases

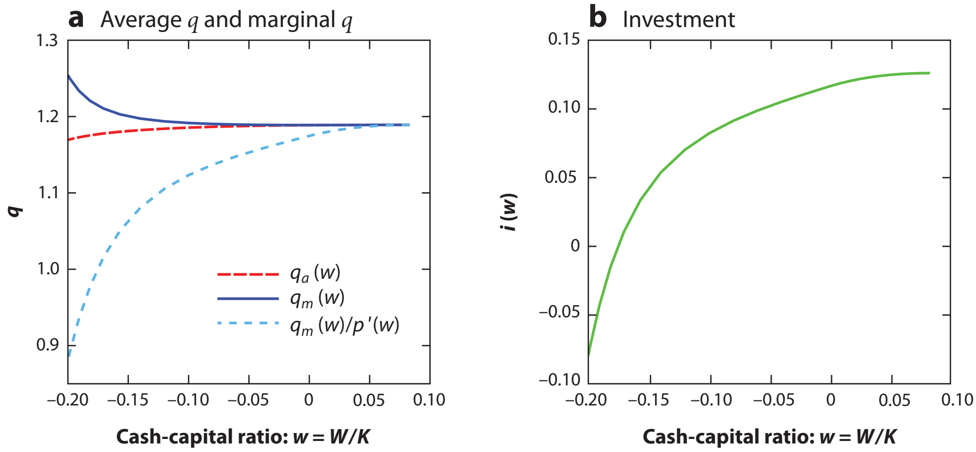

Figure 7

Investment and q with credit line. Panel *a* shows the average q, $q_a(w)$, marginal q, $q_m(w)$, and the ratio of marginal q to the marginal value of liquidity, $q_m(w)/p'(w)$. Panel *b* shows the investment-capital ratio, $i(w)$. Figure adapted from Bolton, Chen & Wang (2011).

with w throughout the credit region. Investment, however, is governed by the ratio of marginal q and the marginal value of cash $p'(w)$, which is also at its highest when w is close to the credit limit. All in all, our model starkly illustrates why marginal q is not a good guide for investment for indebted firms even when putting the measurement issues aside.

Note also that marginal q can be higher than the first-best marginal q^{FB} in the credit region. This certainly does not mean that a highly financially constrained firm should invest more than under the first-best. To the contrary, investment is at its lowest level at $w = -c = -0.2$. This is because the marginal value of cash $p'(w)$ is also at its highest level when $w = -c$. Thus, a key takeaway from this analysis is that marginal q and the marginal value of cash may be highly correlated.

An empirical observation is that small firms rarely access the public equity market but instead largely manage their funding needs through credit lines. One way to interpret this empirical observation via the lens of our model is as follows. In practice, the fixed costs of external financing have a component independent of firm size (see, e.g., Altinkilic & Hansen 2000). As argued by Myers & Majluf (1984), small firms also tend to be more opaque and thus face higher costs when issuing information-sensitive securities.[13] In our model, both of these considerations imply a larger ϕ_0 for small firms than for large firms. With this interpretation, our model then predicts that smaller firms are less willing to raise external equity and depend more on internal cash and credit line than large firms, *ceteris paribus*. Indeed, when firm size is so small (causing ϕ_0 to be sufficiently large), our model predicts that this firm prefers not to issue any external equity, consistent with the evidence above.

3.2.4. Hedging. A firm facing external financing costs is endogenously risk averse with respect to unpredictable changes in corporate savings w. This aversion to risk is reflected in $-p''(w)/p'(w)$,

[13] Bolton, Wang & Yang (2023) develop a dynamic trade-off model in which a firm facing costly external equity financing preserves its financial flexibility by keeping its leverage low in a jump-diffusion model. Paradoxically, a low target leverage is not due to debt being more costly, but rather to the firm's future needs to raise costly equity to service its high level of debt should it find itself in financial distress.

the curvature of the value function. It follows that a firm facing external financing costs may benefit from hedging the risk in w by using state-contingent contracts, such as futures and derivatives contracts. This benefit is over and above any benefits that shareholders can obtain directly by hedging and diversifying their own portfolio risk.

Bolton, Chen & Wang (2011) show that firm value $p(w)$ increases when the firm takes futures positions that reduce its cash flow risk. Furthermore, the demand for precautionary savings is reduced and the firm pays out retained earnings earlier. Interestingly, the effects of cash flow hedging on investment are highly nonlinear. When w is very low, the firm invests even less when it engages in hedging. This is a surprising result. Indeed, in a static model, hedging generally mitigates underinvestment, as Froot, Scharfstein & Stein (1993) have shown. But in a dynamic setting, another logic sets in because asset sales can also be used as a risk management device. By selling assets more aggressively when w is low, the firm increases its going-concern value. When a firm engages in hedging, its going-concern value is higher. Therefore, the benefits from engaging in asset sales when w is low are greater. When the firm faces severe financing constraints, it is optimal to aggressively scale back operations in order to preserve cash and thus be able to better manage risk. In sum, hedging and asset sales are complementary risk management tools in a dynamic setting.

4. CORPORATE SAVINGS AND ASSET PRICING

The presence of financial constraints can amplify a firm's sensitivity to various shocks. Aggregate financing shocks can also be a source of priced risks. As a result, firm beta and expected returns will be tightly connected to financial constraints, for which the marginal value of cash is again a key determinant. In this section, we examine some of the asset pricing implications of financing constraints and the related literature.

First, in the presence of financial constraints, the standard textbook formula for delevering the beta for firms that hold cash is no longer valid. Second, in contrast to the conventional wisdom, idiosyncratic volatility will affect expected returns through its effects on the marginal value of cash and firm beta. Third, empirical studies of the relationship between cash holding and returns need to take into account the endogeneity of cash holdings. Finally, we discuss how aggregate shocks to financing conditions affect asset pricing.

In a conditional CAPM framework, Bolton, Chen & Wang (2011)[14] show that a firm's conditional market beta in the presence of financing friction is

$$\beta(w) = \frac{\rho\sigma}{\sigma_m}\frac{p'(w)}{p(w)}, \qquad 28.$$

where ρ is the correlation between the firm's productivity shock dA_t and returns of the market portfolio, σ is the volatility of dA_t, and σ_m is the volatility of the market portfolio. In contrast, the beta of an unconstrained firm is constant and is given by

$$\beta^{FB} = \frac{\rho\sigma}{\sigma_m}\frac{1}{q^{FB}}, \qquad 29.$$

where q^{FB} is given in Equation 16.

Equation 28 demonstrates the role of the marginal value of cash $p'(w)$ in firm beta. All else equal, a higher $p'(w)$ means higher sensitivity of firm value to cash flow shocks, which translates into a higher beta. Furthermore, a tightening of financial constraint, either due to a decline in the

[14]For more information, see Bolton, Chen & Wang's (2011) internet appendix, section A.

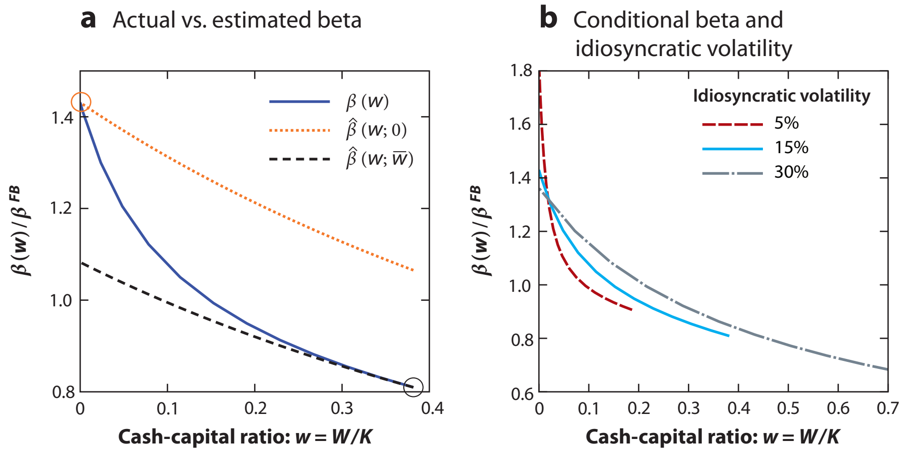

Figure 8

Cash holding, idiosyncratic volatility, and market beta. In panel *a*, we compare the actual market beta as a function of the cash-capital ratio $\beta(w)$ against the estimated beta $\widehat{\beta}(w;\, w_0)$ based on the textbook unlevered beta formula, with $w_0 = 0.3$ (denoted by the *orange circle*). In panel *b*, we plot $\beta(w)$ for three different levels of idiosyncratic volatility (5%, 15%, 30%). The right end of each line corresponds to the respective payout boundary. The black circle indicates the point where the blue solid line intersects with the black dashed line.

firm's cash-capital ratio w or an exogenous increase in external financing costs (e.g., the fixed cost ϕ), will raise the marginal value of cash $p'(w)$ and lower the firm value $p(w)$, both of which push the conditional beta $\beta(w)$ higher.

4.1. Cash and Levered Beta

Let us now contrast Equation 28 with the standard textbook treatment on how to unlever the beta for firms with cash and debt (see, e.g., Berk & DeMarzo 2019). When ignoring financial constraints, one can treat cash as negative debt and view a firm as simply a portfolio of debt/cash and equity. This would imply the unlevered beta is a weighted average of the beta for equity and net debt. Applying this law-of-one-price logic, and by setting the beta of cash to zero, we can start with a firm with any given cash-capital ratio $w_0 \geq 0$ and equity beta $\beta(w_0)$, which could be estimated from the data, and calculate the implied equity beta when the cash holding is changed from w_0 to w,

$$\widehat{\beta}(w;\, w_0) = \frac{p(w_0)}{p(w_0) + (w - w_0)}\beta(w_0). \qquad 30.$$

Figure 8*a* illustrates how sizable the approximation error for the conditional beta can be when ignoring financial constraints. We assume that one can estimate the actual beta from the data, either when $w_0 = 0$ or when $w_0 = \overline{w}$ (at the payout boundary), and then estimate beta at different levels of cash-capital ratio w according to the formula in Equation 30. The solid line in **Figure 8*a*** shows the actual beta $\beta(w)$, while the dotted and dashed lines show the estimated beta $\widehat{\beta}(w;\, w_0)$ for the two different choices of w_0. All of the betas are normalized by β^{FB}, the beta of an unconstrained firm.

When w is near the payout boundary $\overline{w}$, the firm is close to being unconstrained. Thus, it is not surprising that, when starting from $w_0 = \overline{w}$, the estimated beta is close to the actual beta when w is high. However, as w drops lower, not only the relative weights of cash and physical capital on the balance sheet change but also the financial constraint tightens. As a result, the estimated beta increasingly underestimates the actual beta (see the black dashed line). When $w = 0$, the estimated beta is 24% lower than the actual beta.

The opposite is true when we start with a cash-poor firm with $w_0 = 0$. The formula in Equation 30 will overestimate the firm's beta under higher cash holding (see the red dotted line), as it ignores the fact that higher cash holding on the balance sheet has the additional effect of reducing the sensitivity of firm value to cash flow shocks. When w rises to $\overline{w}$, the estimated beta is 32% higher than the actual value.

Notice also that, when w is sufficiently high, the beta for a firm facing financing frictions can become even lower than that of the unconstrained firm (with the blue solid line in **Figure 8*a*** dropping below 1). For example, at the payout boundary, the conditional beta can be as low as 81% of the first-best beta. We may illustrate this point by rewriting the conditional beta as follows:

$$\beta(w) = \frac{\rho\sigma}{\sigma_m} \frac{p'(w)}{(p(w) - w) + w} = \frac{\rho\sigma}{\sigma_m} \frac{p'(w)}{q_a(w) + w}, \qquad 31.$$

where $q_a(w) = p(w) - w$ is the firm's average q (the ratio of the firm's enterprise value and its capital stock). Although $q_a(w) < q^{FB}$ and $p'(w) > 1$, the second term, w, in the denominator of $\beta(w)$ can be so large that $\beta(w) < \beta^{FB}$. Intuitively, as a financially constrained firm hoards cash to reduce external financing costs, the firm beta becomes a weighted average of the asset beta and the beta of cash (zero). With a large enough buffer stock of cash holdings relative to its assets, this firm can become even safer, as measured by market beta or risk premium, than neoclassical firms facing no financing costs and holding no cash.

4.2. Beta and Idiosyncratic Volatility

A common perception about idiosyncratic risk is that it can be diversified away and thus should not affect a firm's expected return. However, for a financially constrained firm, a rise in the idiosyncratic volatility of cash flows can affect the firm's expected return by changing its beta on priced shocks. Here, we again use the Bolton, Chen & Wang (2011) model to demonstrate the link between idiosyncratic cash flow volatility and market beta.

Consider two firms with identical balance sheets, in particular, the same cash-capital ratio, but one firm has a higher level of idiosyncratic cash flow volatility than the other. As Equation 28 shows, the firms' market betas depend on their marginal value of cash $p'(w)$, which is affected by idiosyncratic volatility in two ways. First, when w is sufficiently high, the low-volatility firm will behave like it is almost unconstrained [with $p'(w)$ close to 1], while the high-volatility firm faces a higher probability of running out of cash and thus has a higher $p'(w)$. This force suggests that the marginal value of cash and market beta will increase in idiosyncratic cash flow volatility when w is not too low.

Second, the low-volatility firm also has a higher continuation value, because it is less likely to end up in a state of financial distress. This means that as w approaches zero, its marginal value of cash will rise faster than that of the high-volatility firm and will eventually rise to a higher level. This force suggests that the marginal value of cash and market beta will increase in idiosyncratic cash flow volatility when w is sufficiently close to zero.

Combining these two effects, we get a state-dependent relation between market beta and idiosyncratic volatility as shown in **Figure 8*b***. At low levels of w, the normalized beta $\beta(w)/\beta_{FB}$ can approach as high as 1.8 for the firm with idiosyncratic volatility of 5%, but the corresponding peak normalized beta is less than 1.4 for the firm with 30% idiosyncratic volatility. On the other hand, when we are sufficiently away from the left boundary (which applies most of the time), the market beta is rising with idiosyncratic volatility when holding w fixed.

4.3. Cash and Beta in the Cross Section

Besides demonstrating how idiosyncratic volatility can affect market beta, **Figure 8*b*** also helps us think about the cross-sectional relationship between cash holding and stock returns.[15] When we treat the cash holding of a firm as exogenous, all else equal, firms with higher w will have lower market betas and lower expected returns. Next, let us examine the effect of the endogeneity of cash holding by considering two firms facing different levels of idiosyncratic cash flow volatility. The high-volatility firm will endogenously choose a higher payout boundary, and its average cash holding will be higher than that of the low-volatility firm. However, the market beta of the high-volatility firm could still be higher than that of the low-volatility firm, due to the fact that higher idiosyncratic volatility tends to raise market beta when w is not too low. As a result, in the cross section, a firm that chooses to have a high cash holding could have higher expected returns than one with low cash holding.

The positive relation between expected returns and cash holdings will further strengthen if the heterogeneity is due to systematic rather than idiosyncratic cash flow volatility. Palazzo (2012) argues that firms whose cash flows are more exposed to aggregate shocks are likely to have a stronger precautionary savings motive, which would imply a positive relation between expected equity returns and cash holdings. Moreover, his model implies that this positive relation should be stronger for firms with less valuable growth options, because the riskiness of such firms is more closely tied to the cash flow risks of their assets in place. Empirically, he shows that when sorting stocks into deciles based on the cash-to-total assets ratio, the top decile (with highest cash holdings) earns an excess return of 0.69% per month (equal-weighted) over the bottom decile. Moreover, he finds that firms with higher expected returns (using an accounting-based ex ante measure) tend to see a larger subsequent increase in cash holding.

The endogeneity of cash holding and its impact on the cross section of asset prices also applies to the pricing of corporate bonds. Acharya, Davydenko & Strebulaev (2012) show that the correlation between cash holding and credit spreads is robustly positive. They also explain this fact through the endogeneity of cash holding due to precautionary motives. Riskier firms choose to hold more cash, which might reduce default risk in the short-term, but the longer-term probability of default will still be higher, resulting in higher credit spreads for longer-term bonds (with time-to-maturity exceeding 1 year).

One strategy to address the endogeneity issue with cash holding is by examining the impact of large unanticipated shocks on firms that are caught with different levels of internal financial slack. For example, Duchin, Ozbas & Sensoy (2010) find that a portfolio of cash-rich firms (based on the cash balances at the end of 2006) outperforms a cash-poor portfolio during the onset of the GFC (second half of 2007). Ramelli & Wagner (2020) find that, even within industries and after controlling for standard firm characteristics, firms with low cash holdings at the end of 2019 had more negative stock returns during the early periods of the COVID-19 pandemic (from January to March 2020).

We conclude this section by discussing how aggregate financing shocks can affect the pricing of financially constrained firms within the Bolton, Chen & Wang (2011) framework. Bolton, Chen & Wang (2013) extend the 2011 framework by allowing the aggregate financing condition to fluctuate over time following a Markov chain. In the presence of aggregate financing shocks, the

[15]The studies we discuss here are part of a bigger literature on the relation between financial constraints and stock returns. See, e.g., Lamont, Polk & Saa-Requejo (2001), Whited & Wu (2006), and Livdan, Sapriza & Zhang (2009), among others.

conditional risk premium is determined by a two-factor model, which prices both the systematic cash flow shocks and the shocks to financing condition. In that setting, Equation 28 still determines the exposure of firm equity to systematic cash flow shocks within a given financing state, although both the firm value-capital ratio $p(w)$ and the marginal value of cash $p'(w)$ will become state-dependent. In addition, there is a new financing premium component, which is the product of the intensity of financing shocks, the relative jump in the stochastic discount factor upon the financing shock, and the relative drop in firm value upon the shock (for more details, see section 6 of Bolton, Chen & Wang 2013). For more extensive analysis of the implications of financing shocks on asset pricing, see Belo, Lin & Yang (2019).

5. THE MARGINAL VALUE OF CASH: NEW EXPLORATIONS

The main idea of the dynamic analytical framework developed in Section 3 is that it is optimal for a firm to manage its liquidity and risk when facing costly external financing. This is in effect a theorem of corporate finance, because the alternative of not saving inside the firm is suboptimal—as, without savings, the firm is forced to access external capital markets too often, which is too costly.[16]

This article is not meant to provide a comprehensive survey of the related literature. Instead, we use this analytical framework simply as a starting point to highlight the key insights. This simple framework leaves many important questions unanswered.

For example, what if financing conditions change over time? Should the firm time the market? What if investment is lumpy, costly to reverse or simply irreversible? In the framework developed in Section 3, external financing costs are exogenously given. But external financing costs arise endogenously, possibly due to informational asymmetry, agency problems, or inalienable human capital. How do these frictions drive a firm's liquidity and risk management policies? How does a multidivision firm manage its liquidity and risk management policies? How do the bright and dark sides of internal capital markets determine a multidivision firm valuation and policies?

5.1. Financial Shocks and Market Timing

Another plausible cause for a substantial increase in corporate savings is the anticipation of a potential future financial crisis, which causes a jump in the cost of external financing and possibly even a financial market shutdown. The 2008 GFC exemplifies significant uncertainties in corporate financing conditions. Generalizing the Bolton, Chen & Wang (2011) framework, Bolton, Chen & Wang (2013) develop a dynamic q-theoretic framework where firms have both a precautionary-savings motive and a market-timing motive for external financing and payout decisions, induced by stochastic financing conditions.

During a crisis like the GFC, the cost of external financing can become substantially higher (it could be so high that external financing is effectively unavailable). In response, firms delay payout, cut investment, and engage in fire sales of assets even when their productivity remains unaffected, all to avoid incurring prohibitive equity issuance costs. This is especially true when a firm enters the crisis with low cash reserves. Such heterogeneity in firm responses to a financial shock is

[16]For early inventory models applied to corporate finance, see Baumol (1952), Tobin (1956), and Miller & Orr (1966). The marginal value of liquidity also plays a critical role in international finance and public finance. Rebelo, Wang & Yang (2022) develop a sovereign-debt model under limited financial development in which the marginal cost of servicing sovereign debt plays a critical role. Jiang et al. (2024) develop a p theory of taxes and debt management with an endogenous marginal cost of servicing government debt, debt limit, and debt-GDP transition dynamics.

consistent with empirical and survey evidence of Campello, Graham & Harvey (2010), Duchin, Ozbas & Sensoy (2010), and Campello et al. (2011), among others, and can help us distinguish between financial and productivity shocks empirically.

When external financing costs are affordable (cheap), firms time their equity offerings and issue equity even when there is no immediate need for external funds.[17] Just as firms with low cash holdings seek to take advantage of low costs of external financing to raise more funds, firms with high cash holdings are inclined to disburse their cash through stock repurchases when financing conditions improve. This result is consistent with the finding that aggregate equity issuance and stock repurchases are positively correlated. When the perceived probability of a crisis rises, firms invest more conservatively, issue equity sooner, and delay payouts to shareholders, all to increase cash hoards that will help them through the impending crisis.

One important implication of the model for empirical research is that it can be misleading to assess the impact of financial shocks entirely based on ex post responses such as declines in investment and output. The real effects of financing shocks on a firm crucially depend on the probability that the firm attaches to the financing shock, which in turn affects the distribution of firm cash holding when the shock arrives. A relatively small rise in the probability of a financing shock can already cause firms to underinvest significantly and hold onto excess cash in good times, which in turn leads to small average investment responses to financing shocks ex post.

It is also worth noting that an important and active area of research in macro-finance is in understanding the financial and real impact of large fluctuations in credit conditions, with financial crisis being a prime example. Jermann & Quadrini (2012) build a general equilibrium model in which financial shocks are modeled as stochastic variations in the tightness of a debt enforcement constraint. By exploiting model restrictions, they extract a time series of the shock series from empirical data for debt, capital, and output, and they find that financial shocks help explain the cyclical patterns in firm financial flows and real business cycle quantities. However, for tractability, most models in this literature do not allow for active cash saving decisions. Enriching the models along this dimension will be a worthwhile effort.

While Bolton, Chen & Wang (2013) focus on time-varying external equity financing costs, the literature has also studied other forms of fluctuations in financing conditions. For example, Hugonnier, Malamud & Morellec (2015a) develop a dynamic model of investment, financing, and cash management decisions in which investment is lumpy and firms face capital supply uncertainty. Specifically, they model external financing opportunities using a Poisson process.[18] Hugonnier, Malamud & Morellec (2015b) examine the implications of uncertainties about access to credit markets on the capital structure dynamics.

Another important form of uncertainty about financing conditions is firms' access to credit lines (for evidence during the GFC, see, for example, Ivashina & Scharfstein 2010, Acharya & Mora 2015). This consideration has important implications for how firms use their credit lines versus other sources of liquidity. If a firm's demand for liquidity tends to spike when banks experience significant deposit outflows, banks will find it costly to provide a credit line to the firm. Indeed, Acharya, Almeida & Campello (2013) find that firms with higher aggregate risk exposures opt to hold more cash (and be less dependent on credit lines) because it is more costly for them

[17]This is supported by the empirical findings of McLean (2011) and Acharya, Byoun & Xu (2020). In particular, Acharya, Byoun & Xu (2020) show that when the cost of equity capital is cheap, a firm finds it beneficial to build up financial slack and then use it for investments when the cost of equity capital is high.

[18]A nice technical contribution of Hugonnier, Malamud & Morellec (2015a) is that they demonstrate that smooth-pasting conditions may not guarantee optimality and show that firms may not follow standard single-barrier policies.

to obtain credit lines from banks. In the aggregate, a strong correlation between credit line drawdowns and deposit withdrawals means that banks may not be able to reliably serve as liquidity providers for firms in bad times (Kashyap, Rajan & Stein 2002). A thorough examination of credit line supply and demand dynamics over business cycles and their general equilibrium implications is warranted.

5.2. Real Options

In practice, corporate investment is often lumpy and irreversible. This is true for nearly all investments in property development, factories, and productive equipment. The q theory of investment is an imperfect description of such investments, as it is fundamentally a theory of continuous investment (or divestment). The theory of investment based on real options (McDonald & Siegel 1986) is a better model to apply to such investments. Most of this theory, however, assumes that the firm faces no external financing costs and that the Modigliani-Miller theorem holds.

Bolton, Wang & Yang (2019a) extend the classical real options theory by adding external financing costs. This assumption fundamentally changes the basic predictions of real options theory and adds considerable realism to the classical real options model. A firm with low corporate savings delays investment much more than predicted by the classical theory. The reason is again that the firm prefers to fund its capital expenditures with internal funds. The model analyzed in Bolton, Wang & Yang (2019a) is more complex than the model described in Section 3. The reason is that the classical real options model (where the firm faces persistent productivity shocks) augmented with external financing costs is fundamentally a two-dimensional model with a partial differential equation for firm value rather than an ODE. A fundamental prediction for investment dynamics of this model is that when the firm's cash holdings are depleted—following a crisis or a recession—low investment persists even when earnings fundamentals fully recover. To be sure, in the recovery phase following the crisis, firms first and foremost seek to rebuild their corporate savings. It is only when their financial health is sufficiently restored that they are ready to contemplate new investments.

In the framework developed in Section 3, external financing is costly by assumption. Next, we turn to two formulations that provide microfoundations of costly external financing used in our framework.

5.3. Dynamic Agency (Moral Hazard) and Financial Constraints

One foundation for costly external financing is informational friction and/or moral hazard. Extending the optimal financial contracting model of DeMarzo & Sannikov (2006) to analyze corporate investment in a neoclassic q-theoretic framework,[19] DeMarzo et al. (2012) show that incentive contracting generates a history-dependent wedge between marginal and average q, and both vary over time as good (bad) performance relaxes (tightens) financing constraints. One key difference between Bolton, Chen & Wang (2011) and DeMarzo et al. (2012) is the state variable measuring financial slack: cash and credit in the former, and the manager's continuation payoff in the latter. In the former, the manager directly influences the drift but not the volatility of financial slack. In the latter, the manager's equilibrium effort choice affects the volatility of financial slack but not the drift. As a result, investment is monotonically linked to marginal q in the latter, while investment is linked to the ratio between marginal q and the endogenous marginal value of cash

[19]DeMarzo & Sannikov (2006) build on the two-period agency model of Bolton & Scharfstein (1990) and the discrete-time multiperiod formulation of DeMarzo & Fishman (2007a,b). The DeMarzo-Fishman-Sannikov line of work provides an operational paradigm to tractably analyze dynamic agency.

in the former. This distinction leads to different implications for corporate investment, financing policies, and payout to investors in the two models.

In sum, both transaction-cost-based liquidity management models and agency models are useful to understand financial constraints. On the one hand, the transaction-cost-based model developed in Section 3 has a strong microfoundation based on dynamic agency, and its many predictions line up with agency-based model predictions. On the other hand, models with exogenously specified transaction financing costs also generate new predictions absent in dynamic agency models.

5.4. Inalienable Human Capital and Financial Constraints

A common foundation for the presence of financial constraints is the limited ability of the firm to commit to servicing its liabilities, given that human capital is inalienable (Hart & Moore 1994). The model we have described above can be modified to include a participation constraint for the founder of the firm (or other key employees). Such a model is analyzed in Bolton, Wang & Yang (2019b). To ensure continued future participation of key employees, the firm must offer sufficiently attractive future compensation promises, and these promises must be credible. Corporate savings play an important role in giving credibility to the firm's future compensation promises. This is an important reason why tech companies hold so much cash. Corporate risk management in such a situation also plays a special function. It allows the company to offer smoother compensation and to limit the risk exposure of risk-averse employees.

5.5. Cash and Intangible Assets

Falato, Kadyrzhanova & Sim (2022) study a discrete-time version of a q-theoretic model of internal financing that features two types of productive assets, physical and intangible capital. A key new feature is that only tangible capital can serve as collateral. As such, a higher share of intangible capital reduces a firm's debt capacity and makes it more dependent on internal liquidity. This mechanism allows the authors to connect the rising share of intangible capital and the secular rise in corporate cash holdings in the US data.

Dou & Ji (2021) develop an industry equilibrium model of monopolistic competition that links markups to the endogenous corporate liquidity in the presence of financial constraints. Their model predicts that firms' markups will be more responsive to changes in their marginal value of cash when the marginal value of customer base is higher.

5.6. Taxes and Term Debt

In Bolton, Chen & Wang (2014), we add equity issuance costs to the standard dynamic trade-off theory model of capital structure. An important additional cost of debt financing in this expanded model is debt service: Debt payments drain the firm's cash holdings, which increases the risk of incurring equity issuance costs. Also, realized earnings are separated in time from payouts to shareholders, so that savings have both a corporate tax component, when savings are inside the firm, and a personal tax component, when savings are outside the firm. In this setting, standard measures of the net tax benefits of debt are no longer valid.

Hennessy & Whited (2005) also discuss the implications of retained cash for the trade-off theory. They argue that the corporation will want to reduce savings as long as the corporate tax rate τ_c exceeds the tax rate on interest income τ_i, which, as Bolton, Chen & Wang (2014) show, does not necessarily hold in the presence of financial constraints.

Due to the introduction of term debt, another difference of Bolton, Chen & Wang (2014) from the standard literature (see, for example, Hennessy & Whited 2005, 2007) that only considers one-period debt is that it is no longer valid to consider cash as negative debt. The model of Bolton, Chen & Wang (2014) features the coexistence of cash and long-term debt, and it could be optimal for a firm to simultaneously increase its long-term debt and cash holding in anticipation of future liquidity or investment needs. Eisfeldt & Muir (2016) study this question in a richer setting. Their calibrated dynamic model can account for aggregate external financing and savings waves.

5.7. Internal Capital Markets

Generalizing the Bolton, Chen & Wang (2011) framework to account for internal capital markets for a two-division firm, Dai et al. (2024) develop a tractable model of dynamic investment, spin-offs, financing, and risk management for a conglomerate facing costly external finance.[20]

5.8. Decreasing Return to Scale

Recent work by Kakhbod et al. (2023) generalizes the A-K technology in Bolton, Chen & Wang (2011) to allow for diminishing returns to scale. This feature generates implications for heterogeneous firm size and cash policies and helps bring the model closer to the data.

5.9. Macro-Financial Implications

Models of financial constraints and liquidity management have been applied in macro-finance to understand financial institutions and households. For example, they are closely connected to dynamic models of intermediary constraints by He & Krishnamurthy (2013) and Brunnermeier & Sannikov (2014), among many others, who show that the balance sheets of the financial intermediaries matter for financial markets and the real economy. Models by Alvarez & Lippi (2009) and Chen, Michaux & Roussanov (2020) consider the liquidity management of households. Alvarez & Lippi (2009) extend the Baumol-Tobin cash inventory model to a dynamic environment that allows for the possibility of withdrawing cash at random times at a low cost, which they use to explain the impact of financial innovation on money demand. Chen, Michaux & Roussanov (2020) show that the precautionary motives of homeowners, who face counter-cyclical idiosyncratic labor income uncertainty, can account for many salient features in the evolution of balance sheets and consumption in the cross section of households over a housing boom-bust episode from 2001 to 2012.

6. CONCLUDING REMARKS

Absent frictions inside and outside the firm, corporate financing and balance sheets are irrelevant for value creation. Consequently, all corporate decisions boil down to NPV maximization, taking risk premium, growth opportunities, and real flexibility/optionality into consideration (Modigliani & Miller 1958). In the neoclassic q theory of investment, firm investment is thus fully determined by the marginal value of capital, i.e., marginal q. However, when facing financing frictions, e.g., adverse selection, managerial agency, limited commitment, and inalienable human capital, just to name a few, firms inevitably have incentives to build a liquidity buffer (savings) to mitigate the negative impact of these frictions. Consequently, firm investment and valuation are determined by not only marginal q but also the marginal value of cash.

[20]These authors formalize the following insights for multidivision firms: (*a*) within-firm resource allocation is based not only on the divisions' productivity—as in winner-picking models—but also on their risk; (*b*) firms may voluntarily spin off productive divisions to increase liquidity; (*c*) diversification can reduce firm value in low-liquidity states, as it increases the cost of a spin-off and hampers liquidity management; (*d*) corporate socialism makes liquidity less valuable; and (*e*) division investment is determined by the ratio between marginal q and marginal value of cash.

As we emphasize in this review, the marginal value of cash is inherently a dynamic concept, as it depends on a firm's current financial slack, its outside options, and the future uncertainties it faces. It highlights that the asset side of the balance sheet (in particular, the asset liquidity structure) matters critically for firms in the presence of financial frictions. This basic observation implies that much of corporate finance and valuation, asset pricing, and macroeconomic models with financial frictions should include the marginal value of cash and the dynamics of corporate savings in their analysis.

Although significant progress has been made in understanding the role of cash in a variety of corporate decisions, many open questions remain. Examples include the determinants of the timing of investment and financing decisions, the strategic value of cash in product market competition, the aggregate and distributional implications of credit market imperfections, and the role of intangible assets in shaping corporate financial policies.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Robin Greenwood and an anonymous reviewer for comments. We also thank Heqing Huang for outstanding research assistance. Neng Wang acknowledges the support from the Risk Management Institute at the National University of Singapore.

LITERATURE CITED

Acharya VV, Almeida H, Campello M. 2013. Aggregate risk and the choice between cash and lines of credit. *J. Finance* 68(5):2059–116

Acharya VV, Byoun S, Xu Z. 2020. *The sensitivity of cash savings to the cost of capital*. NBER Work. Pap. 27517

Acharya VV, Davydenko SA, Strebulaev IA. 2012. Cash holdings and credit risk. *Rev. Financ. Stud.* 25(12):3572–609

Acharya VV, Mora N. 2015. A crisis of banks as liquidity providers. *J. Finance* 70(1):1–43

Acharya VV, Steffen S. 2020. The risk of being a fallen angel and the corporate dash for cash in the midst of COVID. *Rev. Corp. Finance Stud.* 9(3):430–71

Almeida H, Campello M, Cunha I, Weisbach MS. 2014. Corporate liquidity management: a conceptual framework and survey. *Annu. Rev. Financ. Econ.* 6:135–62

Altinkilic O, Hansen RS. 2000. Are there economies of scale in underwriting fees? Evidence of rising external financing costs. *Rev. Financ. Stud.* 13(1):191–218

Alvarez F, Lippi F. 2009. Financial innovation and the transactions demand for cash. *Econometrica* 77(2):363–402

Anderson RW, Carverhill A. 2012. Corporate liquidity and capital structure. *Rev. Financ. Stud.* 25(3):797–837

Baumol WJ. 1952. The transactions demand for cash: an inventory theoretic approach. *Q. J. Econ.* 66:545–56

Belo F, Lin X, Yang F. 2019. External equity financing shocks, financial flows, and asset prices. *Rev. Financ. Stud.* 32(9):3500–43

Berg T, Saunders A, Steffen S. 2021. Trends in corporate borrowing. *Annu. Rev. Financ. Econ.* 13:321–40

Berk J, DeMarzo P. 2019. *Corporate Finance*. Prentice Hall Ser. Finance. Boston: Pearson Educ.

Bolton P, Chen H, Wang N. 2011. A unified theory of Tobin's *q*, corporate investment, financing, and risk management. *J. Finance* 66(5):1545–78

Bolton P, Chen H, Wang N. 2013. Market timing, investment, and risk management. *J. Financ. Econ.* 109(1):40–62

Bolton P, Chen H, Wang N. 2014. *Debt, taxes, and liquidity*. NBER Work. Pap. 20009

Bolton P, Scharfstein DS. 1990. A theory of predation based on agency problems in financial contracting. *Am. Econ. Rev.* 80(1):93–106
Bolton P, Wang N, Yang J. 2019a. Investment under uncertainty with financial constraints. *J. Econ. Theory* 184:104912
Bolton P, Wang N, Yang J. 2019b. Optimal contracting, corporate finance, and valuation with inalienable human capital. *J. Finance* 74(3):1363–429
Bolton P, Wang N, Yang J. 2023. *Leverage dynamics and costly equity issuance*. NBER Work. Pap. 26802
Brunnermeier M, Sannikov Y. 2014. A macroeconomic model with a financial sector. *Am. Econ. Rev.* 104(2):379–421
Campello M, Giambona E, Graham JR, Harvey C. 2011. Liquidity management and corporate investment during a financial crisis. *Rev. Financ. Stud.* 24(6):1944–79
Campello M, Graham JR, Harvey C. 2010. The real effects of financial constraints: evidence from a financial crisis. *J. Financ. Econ.* 97(3):470–87
Chen H, Michaux M, Roussanov N. 2020. Houses as ATMs: mortgage refinancing and macroeconomic uncertainty. *J. Finance* 75(5):1987–2039
Dai M, Giroud X, Jiang W, Wang N. 2024. A q theory of internal capital markets. *J. Finance* 79(2):1147–97
Decamps J, Mariotti T, Rochet J, Villeneuve S. 2011. Free cash flow, issuance costs, and stock prices. *J. Finance* 66:1501–44
DeMarzo PM, Fishman MJ. 2007a. Agency and optimal investment dynamics. *Rev. Financ. Stud.* 20(1):151–88
DeMarzo PM, Fishman MJ. 2007b. Optimal long-term financial contracting. *Rev. Financ. Stud.* 20(6):2079–128
DeMarzo PM, Fishman MJ, He Z, Wang N. 2012. Dynamic agency and the q theory of investment. *J. Finance* 67(6):2295–340
DeMarzo PM, Sannikov Y. 2006. Optimal security design and dynamic capital structure in a continuous-time agency model. *J. Finance* 61(6):2681–724
Denis DJ, Wang L. 2024. Corporate cash holdings. In *Handbook of Corporate Finance*, pp. 223–48. Cheltenham, UK: Edward Elgar Pub.
Dou WW, Ji Y. 2021. External financing and customer capital: a financial theory of markups. *Manag. Sci.* 67(9):5569–85
Duchin R, Ozbas O, Sensoy BA. 2010. Costly external finance, corporate investment, and the subprime mortgage credit crisis. *J. Financ. Econ.* 97(3):418–35
Dumas B. 1991. Super contact and related optimality conditions. *J. Econ. Dyn. Control* 15:675–85
Eisfeldt AL, Muir T. 2016. Aggregate external financing and savings waves. *J. Monet. Econ.* 84:116–33
Falato A, Kadyrzhanova D, Sim JW. 2022. Rising intangible capital, shrinking debt capacity, and the U.S. corporate savings glut. *J. Finance* 77(5):2799–852
Froot K, Scharfstein D, Stein J. 1993. Risk management, coordinating corporate investment, and financing policies. *J. Finance* 48:1629–58
Greenwald DL, Krainer J, Paul P. 2020. *The credit line channel.* Work. Pap. 2020–26, Fed. Reserve Bank San Francisco
Gryglewicz S. 2011. A theory of corporate financial decisions with liquidity and solvency concerns. *J. Financ. Econ.* 99(2):365–84
Hart O, Moore J. 1994. A theory of debt based on the inalienability of human capital. *Q. J. Econ.* 109:841–79
Hayashi F. 1982. Tobin's marginal q and average q: a neoclassical interpretation. *Econometrica* 50:213–24
He Z, Krishnamurthy A. 2013. Intermediary asset pricing. *Am. Econ. Rev.* 103(2):732–70
Hennessy CA, Whited TM. 2005. Debt dynamics. *J. Finance* 60(3):1129–65
Hennessy CA, Whited TM. 2007. How costly is external financing? Evidence from a structural estimation. *J. Finance* 62(4):1705–45
Hugonnier J, Malamud S, Morellec E. 2015a. Capital supply uncertainty, cash holdings, and investment. *Rev. Financ. Stud.* 28(2):391–445
Hugonnier J, Malamud S, Morellec E. 2015b. Credit market frictions and capital structure dynamics. *J. Econ. Theory* 157:1130–58
Ivashina V, Scharfstein D. 2010. Bank lending during the financial crisis of 2008. *J. Financ. Econ.* 97(3):319–38
Jermann U, Quadrini V. 2012. Macroeconomic effects of financial shocks. *Am. Econ. Rev.* 102(1):238–71

Jiang W, Sargent TJ, Wang N, Yang J. 2024. A *p* theory of taxes and debt management. *J. Finance*. Forthcoming
Kakhbod A, Reppen A, Umar T, Xing H. 2023. *Why is cash u-shaped in firm size?* Work. Pap., Univ. Calif., Berkeley
Kaplan SN, Zingales L. 1997. Do investment-cash flow sensitivities provide useful measures of financing constraints? *Q. J. Econ.* 112:169–215
Kashyap AK, Rajan R, Stein JC. 2002. Banks as liquidity providers: an explanation for the co-existence of lending and deposit-taking. *J. Finance* 57(1):33–73
Keynes JM. 1936. *General Theory of Employment, Interest and Money*. London: Palgrave Macmillan
Lamont O, Polk C, Saa-Requejo J. 2001. Financial constraints and stock returns. *Rev. Financ. Stud.* 14(2):529–54
Livdan D, Sapriza H, Zhang L. 2009. Financially constrained stock returns. *J. Finance* 64(4):1827–62
McDonald R, Siegel D. 1986. The value of waiting to invest. *Q. J. Econ.* 101:707–28
McLean RD. 2011. Share issuance and cash savings. *J. Financ. Econ.* 99(1):146–63
Miller MH, Orr D. 1966. A model of demand for money by firms. *Q. J. Econ.* 80:413–35
Modigliani F, Miller MH. 1958. The cost of capital, corporation finance and the theory of investment. *Am. Econ. Rev.* 48(3):261–97
Myers SC, Majluf NS. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *J. Financ. Econ.* 13(2):187–221
Palazzo B. 2012. Cash holdings, risk, and expected returns. *J. Financ. Econ.* 104(1):162–85
Ramelli S, Wagner AF. 2020. Feverish stock price reactions to COVID-19. *Rev. Corp. Finance Stud.* 9(3):622–55
Rebelo S, Wang N, Yang J. 2022. Rare disasters, financial development, and sovereign debt. *J. Finance* 77(5):2719–64
Riddick LA, Whited TM. 2009. The corporate propensity to save. *J. Finance* 64(4):1729–66
Strebulaev IA, Whited T. 2012. Dynamic models and structural estimation in corporate finance. *Found. Trends Finance* 6(1–2):1–163
Sufi A. 2009. Bank lines of credit in corporate finance: an empirical analysis. *Rev. Financ. Stud.* 22(3):1057–88
Tobin J. 1956. The interest-elasticity of transactions demand for cash. *Rev. Econ. Stat.* 38:241–47
Whited TM, Wu G. 2006. Financial constraints risk. *Rev. Financ. Stud.* 19(2):531–59