

Research Reports

Kansantaloustieteen laitoksen tutkimuksia, Nro 97:2003

data, citation and similar papers at core.ac.uk

brought to you by

provided by Helsingin yliopiston digita

SAMU PEURA

Essays on Corporate Hedging

ISBN 952-10-0697-8 (nid.)

ISBN 952-10-0696-X (pdf)

*'I used to do a little
but a little wouldn't do
so the little got more and more'*

W. A. Rose

Acknowledgements

I would like to express my gratitude to my supervisor, Professor Vesa Kanninen, for his instruction, advice, and constructive criticism at times when it was needed.

Learning at best is a social event. I thankfully acknowledge the cooperation with Ph.D. Esa Jokivuolle on essays three and four, and the cooperation with Professor Jussi Keppo on essay two. I have greatly benefited from discussions and exchanges of ideas with these two scientists also outside the joint projects.

I have received valuable comments and insights from my official examiners, Professor Luis Alvarez and Professor Juuso Välimäki. The comments have led to numerous improvements in the essays.

I have had the pleasure to work in an environment where my doctoral studies have never been considered a liability. My work at Sampo Bank and at its predecessors has provided me with research ideas, some of which have materialized as part of this thesis. I am in special gratitude to Ph.D. Juha Reivonen.

I thank my parents Leena and Tapio, my brother Juha, and Henna-Kaisa Wigren for supporting this undertaking in many ways.

I gratefully acknowledge financial support from Liikesivistysrahasto.

Helsinki, April 2003

Samu Peura

Contents

Essays on Corporate Hedging: an overview	1
Essay 1: 'Dividends, costly external capital, and firm value: the case of constant scale'	12
Essay 2: 'Optimal bank capital with costly recapitalization'	46
Essay 3: 'A Value-at-Risk approach to bank capital buffers: an application to the new Basel Accord'	98
Essay 4: 'A structural model of risky debt with stochastic collateral'	119
Essay 5: 'On the markup interpretation of optimal stopping rules'	145

Essays on Corporate Hedging: an overview

1 Corporate hedging

Corporate hedging should be understood far more broadly than as the use of financial derivatives to hedge specific risks in the firm's cash flow. Concern for risks is reflected in investment and portfolio decisions, financing decisions, decisions on payout policies, as well as in the decision on whether to use financial derivatives or not. This is not to say that risk management subsumes all corporate decision making. Rather, the modern perspective on corporate hedging acknowledges that the amount of risk and the capacity to bear risks can be influenced through operating, financing, as well as explicit risk management decisions. Determining where risks are managed most efficiently is a significant part of the corporate hedging decision.

The fundamental risk choice for a firm is the selection of its operating policy. Were there no financial market imperfections, operating policy could be determined without a special concern for risks, and financing and risk management decisions would follow from the choice of operating policy. This separation is implied by the Modigliani-Miller (MM) theorems (1958, 1961). A corresponding separation takes place when there is no operational flexibility, say because the operating policy is fixed by external constraints. In both cases, financing and risk management policies are determined as optimal reactions to the given operating policy.

Given a fixed operating policy, a corporate may hedge through all of the following financial decisions (this is by no means a complete list): a financial risk management policy which smoothens fluctuations in the firm's earnings (Froot et al., 1993); a debt choice which takes into account earnings risks and the expected costs from financial distress (Leland, 1994); a choice of dividend policy which delays dividends and accumulates buffer stocks of capital to control for the probability of bankruptcy (Milne and Robertson, 1996). These examples are to emphasize that even when the operating policy is taken as given, risk management with financial derivatives is only one way to hedge. Capital structure choice, payout policies, and liquidity management matter alike. In many cases, some of these non-explicit risk management tools may be more effective than risk management through financial

derivatives, and in other cases financial derivatives may not be available at favorable terms to the risk management needs of the particular firm.

In the other extreme, the firm may have considerable operating flexibility, while the firm's financial flexibility may be very limited due to severe financial market imperfections. Then the choice of operating policy will be largely dictated by financial constraints, and corporate hedging will have a very operational content. In particular, many strategic decisions can be thought to take place between several mutually exclusive operating policies that are each associated with a distinct combination of risk and return. The concern for volatility induced by the financial constraints will be reflected in the choice of operating policy. An example of this type of corporate hedging analysis is Radner and Shepp (1996).

Arguably, most firms will be located between these two extremes, possessing some financial as well as operational flexibility. Financial constraints will be reflected in the choice of operating policy, while constraints on operational adjustment will impose restrictions on feasible financial policies. Whether risks are managed through operational adjustment or financial adjustment will ultimately depend on the degree of financial flexibility in contrast to operational flexibility.

2 Literature on corporate hedging

Corporate hedging in the presence of linear investor utility is usually motivated by some type of financial market imperfections. The modern corporate hedging literature is reviewed in the following, organized according to the type of market imperfection.

Smith and Stulz (1985) show how taxes create a motive for corporate hedging. When the effective tax function is convex in income, a firm may reduce its expected tax liability by smoothing its income. This argument is essentially an application of Jensen's inequality. Within the same framework, Smith and Stulz show that the optimal degree of hedging depends on the cost of hedging.

Froot et al. (1993) show how costly external capital motivates corporate hedging. The cost of external capital may exceed the cost of internally generated funds due to informational asymmetries. Hedging adds value by ensuring that the firm has sufficient internal funds to take advantage of attractive investment opportunities. In the presence of costly external

capital the firm may underinvest if internally generated funds fall short of profitable investment opportunities. Financial derivatives can be used to shift internally generated funds into those states and times when they are most needed, i.e. when profitable investment projects requiring large capital investments need to be undertaken. This principle determines optimal risk management strategies.

Froot and Stein (1998), building on the model of Froot et al. (1993), show how costly external capital and incompleteness of capital markets influence capital budgeting and capital structure choice in corporations and financial institutions in particular. Non-traded risks need to be priced according to each bank's internal valuation of cash flows. A bank's ability to absorb losses depends on its capital structure. Hedging against non-traded risks can only be accomplished through adjustment of capital structure, i.e. reduction in leverage. The cost of this hedging, the value of lost tax benefits, has to be passed on to the pricing of the risks. Alternatively, when the adjustment of capital structure is very costly, the bank may invest less aggressively in non-traded risks. Hedging of those risks that are traded will increase the bank's tolerance of leverage and adds value since tax benefits of debt may be taken advantage of. Investment choices, capital structure choices, and risk management choices in this theory are jointly and endogenously determined.

A significant synthesis of capital structure theory was achieved by Leland (1994) (also Leland, 1998), who examined the joint determination of corporate capital structure and corporate debt values. Methodologically, Leland's model is an extension of the classical Merton (1974) model of the firm. Optimal capital structure is influenced by firm risk, taxes, and bankruptcy costs, like in the classical theories, but also by the valuation of corporate debt, which in turn depends on the capital structure of the firm. Leland therefore accounts for the endogeneity of debt values within the context of the problem of determining the optimal capital structure. Equity holders may also choose the timing of bankruptcy optimally to maximize the valuation of levered equity, so that operational flexibility is present in this model as well. Leland (1994) essentially brings together three branches of literature: i) the classical literature on optimal capital structure that deals with taxes and bankruptcy costs, ii) the literature on the valuation of risky debt in structural models, initiated by Merton (1974), and iii) the literature on endogenous termination of operations (Brennan and Schwartz, 1985, Dixit, 1989).

Mello and Parsons (2000) analyze optimal hedging strategies in the presence of an exogenous borrowing constraint. Optimal hedges are shown to have the feature that they minimize the variability in the marginal value of the firm's cash balances. They do this by transferring cash from states in which the marginal value of cash balances is low to states where the marginal value of cash balances is at its highest. In the presence of stochastic factors affecting firm profitability, the marginal value of the firm's cash balances is not perfectly correlated with the firm's actual cash balances. Therefore optimal hedging generally differs from hedging the firm's cash flow. The same holds for other commonly used hedging strategies, such as hedging firm value or hedging sales revenues.

Corporate dividend policies and risk choices in the presence of liquidity constraints have been analyzed e.g. by Milne and Robertson (1996), Radner and Shepp (1996), and Hojgaard and Taksar (1999). In these models, firms generate buffer stocks of liquid assets by delaying dividend payments. Hedges can be simultaneously undertaken to control for the drift and the volatility of the firm's cash flow. Optimal dividend policies and risk reduction strategies are jointly determined. They have the feature that most risk reduction takes place when the capital stock is close to the firm's liquidation point, while dividends are only paid once the capital stock exceeds an endogenously determined safety level. When the capital stock is at the dividend barrier, the firm is locally risk neutral and is run at the risk level delivering maximal expected profits.

The models cited above demonstrate that investment, financing, and payout decisions may be driven by the concern to reduce volatility or the firm's sensitivity to volatility, so that there can be a hedging element to each of these decisions. Moreover, hedging considerations may be reflected in these financial and operating decisions even if financial derivatives are present.

Besides the general literature on corporate hedging cited above, there are several research streams which analyze the implementation or construction of hedges in specific contexts. First, the classical literature on the mechanics of hedging for risk (variance) averse agents concentrates on hedging by futures (e.g. Anderson and Danthine, 1981, Stulz, 1984). Corporate hedging of this type is only a part of corporate financial policy. Second, the plain term 'hedging' in the asset pricing literature refers to the process of replication, i.e. the manufacturing of more complex financial payoffs from more simple ones through dynamic

trading based on Merton's (1973) idea. This is a literature that to a large extent takes the hedging objective as given, and as such is detached from the corporate hedging literature that is concerned with the principles that determine the optimal method and degree of hedging.

3 Overview of the essays

The essays in this collection study both financial and operational hedging from a theoretical point of view. Essays 1 to 3 analyze hedges that are implemented through adjustment of financial policies. Essay 1 studies optimal dividend and capital raising policies for a firm which faces an exogenous minimum capital requirement. Essay 2 studies a similar firm but subject to different capital market imperfections. Essay 3 analyzes banks' choice of capital buffers under the proposed new Basel Capital Accord. Essays 4 and 5 analyze hedges which are operational in nature. Essay 4 analyzes the value of collateral as a hedging instrument for banks. Essay 5 studies models where operational flexibility is related to an option to abandon and to restart operations at a given cost. With the exception of essay 3, the essays share a common methodology in that uncertainty is represented by continuous time diffusion processes.

The essays are motivated by pragmatic issues that risk management professionals in corporates and banks are currently facing. In particular, essay 2 is an attempt to test whether current levels of bank capitalization are adequate from the perspective of a fully optimizing model of bank behavior. Essay 3 is strongly motivated by the ongoing changes in banking regulation and the speculated increases in required bank capital. Essay 4 is an attempt to enrich the theoretical literature on defaultable loan pricing by drawing on common bank practices.

Essay 1: Dividends, costly external capital, and firm value: the case of constant scale

This essay studies optimal dividend and capital raising policies for a constant scale firm operating under a minimum capital requirement. The capital requirement states that the firm at all times must have a positive stock of capital to absorb potential losses from the productive activity, if the firm is to continue operating as a going concern. The productive asset is completely illiquid, so that extra capital can not be obtained through liquidating the productive asset. The firm has access to external capital, but raising of external capital is

subject to a proportional cost and can only take place at a bounded rate. The minimum capital requirement, together with these capital market imperfections, induces the firm to maintain buffer stocks of capital. Dividends are paid out of the capital stock, but only to the extent that the optimal buffer stock of capital can be maintained.

The essay extends the model of optimal dividends studied earlier by e.g. Milne and Robertson (1996) and Asmussen and Taksar (1997), by introducing the option to issue new capital. The option to issue capital lowers the optimal buffer stock of capital, identified as the dividend barrier, by an amount that depends on the severity of the capital market imperfections. In the general case, optimal policies are described by two non-zero barriers, such that dividends are paid, at maximal admissible rate, when the capital stock is above an endogenously determined barrier, and capital is issued, again at maximal admissible rate, when the capital stock is below another endogenously determined barrier. The dividend barrier exceeds the capital issue barrier by a margin which depends on the cost of capital issuance. Severe enough capital market imperfections will cause the firm to abstain from issuing external capital altogether. A separate analysis is also provided for all the limiting cases where the capital market imperfections vanish.

The assumption of constant scale abstracts from investment considerations, yielding quite explicit results. Yet it is well known that models with decreasing, and ultimately vanishing, scale returns generate qualitatively similar optimal policies (Radner, 1998, Alvarez, 2001). Moreover, it is shown in an Appendix to this essay that the model can be interpreted as a normalized form of a more general model with positive scale effects and time dependent profit flow.

Essay 2: Optimal bank capital with costly recapitalization

This essay analyzes the same basic model as essay 1, but under different capital market imperfections. Here the issuance of new capital is subject to a fixed cost and an implementation delay. The fixed cost can be interpreted as a fee to an investment bank organizing the capital issue, and the implementation delay represents the time it takes to get access to external capital.

The fixed cost prevents continuous adjustments of the capital stock by means of new capital issues. The optimal capital issuance policy is by nature an impulse control policy, so

that capital is issued at stopping times which are the first-hitting times of endogenously determined barriers. A characterization of the solution in terms of a set of quasi-variational inequalities is provided, and a quasi-analytic solution in terms of two barriers which solve a pair of non-linear equations is obtained. The limiting cases of the model where the capital market imperfections are not present are also solved. The study of the limiting cases reveals that the fixed cost is not the characteristic that drives the qualitative results from the model. In terms of the qualitative nature of the solution, the delay in organizing the capital issue is more fundamental than the fixed cost. This result may have certain methodological value which extends outside the particular interpretation of variables in this model.

The model is calibrated to data on actual bank returns over the period 1994-2001, and is found to be unable to explain the high bank capital levels that are observed empirically. This failure is partly due to banks' accounting options for provisioning of expected losses, which smoothen banks' accounting income. The assumption of normally distributed bank returns is also counterfactual. In the numerical applications, the model is used with implied bank return volatilities analogously to the manner the Black-Scholes formula is used in practice.

Essay 3: A Value-at-Risk approach to bank capital buffers: an application to Basel II

This essay analyzes the determination of bank capital in a more realistic institutional setting compared to essays 1 and 2. It is assumed that bank capital is determined by a Value-at-Risk type criterion which takes into account the volatility in the minimum capital requirement. Monte Carlo simulation techniques are then applied to quantify banks' potential capital holdings under the proposed new Basel Capital Accord (Basel Committee, 2001).

By their design, the minimum capital requirements under the proposed new capital regimes will be sensitive to banks' portfolio quality. A bank's capital charge will therefore vary over time in accordance with its portfolio ratings distribution. Costs associated with capital adjustment and portfolio adjustment create a motive for banks to hedge against shocks to capital. Holding of buffer capital constitutes the natural hedging vehicle, and in fact the only feasible hedging vehicle when the bank's portfolio is highly illiquid and access to external is subject to high costs. Assuming that hedging is based on a targeted confidence level, increased volatility in minimum capital requirements will lead to higher capital buffers in relative terms, i.e. as a percentage of the minimum requirement. Yet the proposed new requirements will simultaneously lower the minimum capital requirements of some banks and

increase those of other banks. Therefore, the effect of the regulatory reform on total capital held by different banks is not immediately obvious. It is shown in this essay that bank capital buffers, as well as banks' total capital holdings, under the new regime are likely to depend in a nontrivial way on the individual banks' portfolio risk and on the selected capital regime.

This simulation based analysis is, to the best of the authors' knowledge, the first quantitative exploration of the implications of risk sensitive bank capital requirements in a stochastic framework and at this level of institutional detail. This is a partial equilibrium analysis, however, which takes a bank's portfolio as given and derives its capital needs based on a choice of risk level by the bank's management. Equilibrium models of the banking sector where capital requirements play a role have been recently studied e.g. by Danielsson et al. (2001). These models abstract from the accurate description of minimum capital rules and bank portfolios that the analysis in this essay is based on.

Essay 4: A structural model of risky debt with stochastic collateral

This essay studies a problem of great practical importance in bank risk management, the evaluation of the protective value of collateral. The Merton (1974) model of risky debt is extended by allowing an explicit collateral value process which is correlated with the asset value process that determines default. The model yields a quasi-analytic expression for the loss-given-default (LGD). The LGD estimate takes into account the key drivers of recovery in the event of bankruptcy: current collateral value, collateral value volatility, and the correlation of collateral value with the default probability of the obligor. LGD estimates are frequently needed in portfolio credit risk models which are run with constant LGD parameters for computational reasons.

As a second application, the model is used to set collateral requirements on bank loans according to a number of criteria which appear to correspond to common bank practices. The analysis is based on the observation that a bank can substitute risk sensitive pricing of its loans with appropriate adjustment of collateral requirements. This is because sufficient protective collateral limits potential losses in the event of default and makes loans homogenous in terms of their riskiness to be eligible for a uniform pricing. Casual evidence suggests that bank loans in a number of countries are priced in this manner.

This essay solves the pattern of collateral requirements implied by the uniform price criterion. Then two alternative criteria for setting collateral requirements are studied which circumvent the problem of pricing the loan in the first place. These criteria are based on limiting the risk from the collateralized exposure in a probabilistic sense. Non-price criteria may have practical value in that the pricing of illiquid bank loans is arguably much greater challenge than the evaluation of the probabilistic behavior of the exposure and the collateral. It is found that one of the probabilistic rules studied yields collateral requirements whose qualitative behavior is closely in line with the uniform price criterion.

Essay 5: On the markup interpretation of optimal stopping rules

This essay studies models where firms' profitability fluctuates stochastically and operational flexibility is present in the form of abandonment and restart options. Firms maximize their value through selection of optimal option exercise policies. Exercising the options is costly, resulting in behavior which is referred to as hysteresis in the real option literature.

The models analyzed here have earlier been studied in Dixit (1989) and in Dixit and Pindyck (1994). The main goal of this essay is to demonstrate, extending on an idea presented by Dixit et al. (1999), that the optimal stopping policies in these models can be characterized through first-order conditions that have the markup property familiar from classical producer theory. These first-order conditions are arrived at when the objective function of the problem is evaluated directly as a function of the unknown stopping barriers. The solution procedure is an alternative to the dynamic programming/variational inequality route. It is claimed that in many cases the direct route, which relies on an appropriate representation of the objective based on renewal arguments, has some clear benefits over the dynamic programming approach. The direct route is also very intuitive and as such may have pedagogic value.

References

- Alvarez, L. H. R., 2001, Singular stochastic control, linear diffusions, and optimal stopping: a class of solvable problems, *SIAM Journal of Control and Optimization*, 39, 6, 1697-1710.
- Anderson, R. W., J. P. Danthine, 1981, Cross hedging, *Journal of Political Economy*, 89, 1182-1196.
- Asmussen, S., M. Taksar, 1997, Controlled diffusion models for optimal dividend payout, *Insurance: Mathematics and Economics*, 20, 1-15.
- Basel Committee on Banking Supervision, 2001, The New Basel Capital Accord, *Consultative Document, Bank for International Settlements*, January.
- Brennan, M. J., E. S. Schwartz, 1985, Evaluating natural resource investments, *Journal of Business*, 58, 135-157.
- Danielsson, J., H. S. Shin, J. P. Zigrand, 2001, Asset price dynamics with Value-At-Risk constrained traders, Financial Markets Group, London School of Economics.
- Dixit, A., 1989, Entry and exit decisions under uncertainty, *Journal of Political Economy*, 97, 620-638.
- Dixit, A., R. Pindyck, 1994, *Investment Under Uncertainty*, Princeton University Press, Princeton, New Jersey.
- Dixit, A., R. Pindyck, S. Sodal, 1999, A markup interpretation of optimal investment rules, *The Economic Journal*, 109, 179-189.
- Froot, K. A., D. A. Scharfstein, J. C. Stein, 1993, Risk management: coordinating corporate investment and financing policies, *The Journal of Finance*, 48, 5, 1629-1658.
- Froot, K. A., J. C. Stein, 1998, Risk management, capital budgeting, and capital structure policy for financial institutions: an integrated approach, *Journal of Financial Economics*, 47, 55-82.
- Hojgaard, B., M. Taksar, 1999, Controlling risk exposure and dividend payout schemes: insurance company example, *Mathematical Finance*, 9, 2, 153-182.
- Leland, H. E., 1994, Corporate debt value, bond covenants, and optimal capital structure, *The Journal of Finance*, 49, 4, 1213-1252.

- Leland, H. E., 1998, Agency costs, risk management, and capital structure, *The Journal of Finance*, 53, 1213-1243.
- Mello, A. S., J. E. Parsons, 2000, Hedging and liquidity, *The Review of Financial Studies*, 13, 1, 127-153.
- Merton, R., 1974, On the pricing of corporate debt: the risk structure of interest rates, *The Journal of Finance*, 29, 449-470.
- Miller, M., F. Modigliani, 1961, Dividend policy, growth, and the valuation of shares, *Journal of Business*, 34, 411-433.
- Milne, A., D. Robertson, 1996, Firm behavior under the threat of liquidation, *Journal of Economic Dynamics and Control*, 20, 1427-1449.
- Modigliani, F., M. Miller, 1958, The cost of capital, corporation finance, and the theory of investment, *American Economic Review*, 48, 261-297.
- Radner, R., 1998, Profit maximization with bankruptcy and variable scale, *Journal of Economic Dynamics and Control*, 22, 849-867.
- Radner, R., L. Shepp, 1996, Risk vs. profit potential: a model for corporate strategy, *Journal of Economic Dynamics and Control*, 20, 1373-1393.
- Smith, C. W., R. Stulz, 1985, The determinants of firms' hedging policies, *The Journal of Financial and Quantitative Analysis*, 20, 391-405.
- Stulz, R., 1984, Optimal hedging policies, *Journal of Financial and Quantitative Analysis*, 19, 127-140.

Essay 1: Dividends, costly external capital, and firm value: the case of constant scale

Abstract

We study optimal dividend and capital raising policies for a constant scale firm operating subject to a minimum capital requirement. Operating profits and losses drive the firm's capital stock. Owners derive value from the stream of dividends distributed out of the capital stock, but the threat of capital shortage constrains dividend distribution. Owners may obtain external capital to reduce the probability of capital shortage, but recapitalization is subject to proportional cost, and can only take place at limited rates. We solve for the optimal policy and the resulting firm valuation, and analyze the effects of costly external capital on firm value. We identify the conditions on the capital market imperfections under which costly recapitalization is optimal, and study the limiting cases of the model as the capital market imperfections vanish. Finally we compare capital issues and operational risk reductions as alternative methods of hedging against capital shortages.

Keywords: dividends, capital issues, firm valuation, proportional cost

JEL classification: G22, G32

1 Introduction

A stream of recent research has studied optimal dividend and risk choices for a constant scale firm which faces an exogenous minimum capital requirement (Jeanblanc-Pique and Shiryaev, 1995, Milne and Robertson, 1996, Radner and Shepp, 1996, Asmussen and Taksar, 1997, Hojgaard and Taksar, 1999, 2001, Asmussen, Hojgaard, and Taksar, 2000). In these models the capital stock in the absence of controls evolves as an arithmetic Brownian motion. Dividends are paid out of the capital stock, and the value of the firm is the present value of optimally chosen dividends. The basic model of dividend optimization in the absence of risk control is treated in Jeanblanc-Pique and Shiryaev (1995), Milne and Robertson (1996) and Asmussen and Taksar (1997). Simultaneous risk choice between a finite number of drift-

volatility coefficient pairs which together satisfy certain monotonicity requirements is analyzed in Radner and Shepp (1996). In Hojgaard and Taksar (1999), the volatility of the diffusion may be scaled down, in which case the drift is scaled down in the same proportion. This risk choice is interpreted as cheap reinsurance for an insurance corporation. Asmussen, Hojgaard and Taksar (2000) analyze a variation of the previous called excess-of-loss reinsurance. In Hojgaard and Taksar (2001), cheap reinsurance is available, and additionally the capital stock earns a possibly stochastic rate of return.

In this paper we extend the basic model of dividend optimization by allowing for costly issues of external equity capital. The value of the firm in our model is the present value of net equity distributions, i.e. the present value of the dividend payout less the present value of the capital issued. We assume that capital issuance carries a proportional cost, and that capital issues can only be implemented at a limited rate. The proportional cost can be interpreted e.g. as the fee to an organizer of the capital issue, and the limited rate reflects the time it takes to obtain external finance. These capital market imperfections create a motive for buffer stocks of capital, the optimal size of which depends on the severity of the imperfections, as well as on the characteristics of the firm's cash flow. We allow dividend payments at unbounded rates.

Both capital issues and risk reductions can be interpreted as risk management tools which allow the firm (a bank or an insurance company) to operate with less buffer capital. There is usually a cost associated with both types of hedges: capital issues come with a proportional cost, while risk reductions as in Radner and Shepp (1996) or Hojgaard and Taksar (1999) involve a partial sacrifice of expected profit. Our model yields a decomposition of the firm value into the value of the capital issue option and the value of the non-optional firm, and enables us to compare the value of capital issues against the value of risk reductions. We illustrate these comparisons in Section 4 of this paper. Our analysis also contributes to the literature that studies optimal corporate hedging mechanisms in the presence of capital market imperfections, in particular in the presence of costly external capital. Froot et al. (1993, 1998) e.g. have shown that costly external capital leads to the interaction between investment and financing, and creates a motive for risk management¹. Our model demonstrates that even in

¹ A related work is also Mello and Parsons (2000), who show how a liquidity constraint generates a motive for risk management and determines the nature of optimal risk management policies.

the absence of investment considerations (our model assumes a fixed portfolio), optimal dividend and capital issuance policies depend on the degree of capital market imperfections associated with capital issuance. Our model does not capture explicit risk management technologies, such as financial derivatives. Risk management takes place by regulating the firm's buffer stock of capital through dividends and capital issues.

The optimal policy in our model, which we obtain by combined regular/singular control methods, is of the barrier control type. In particular, for not-too-extreme parameter values there exists a positive barrier b_1 , such that new capital issues are implemented at maximal admissible rate when the firm's capital stock is below b_1 . The firm starts distributing dividends, again at the maximal admissible rate, when the capital stock is above another barrier b_2 . The difference $b_2 - b_1$ is positive when the cost of capital issues is positive, approaches zero when the cost of capital issues approaches zero, and does not depend on the maximal admissible rate of capital issuance. Moreover, as the cost of capital issues exceeds a critical barrier, it is no more optimal for the firm to resort to capital issues. In this case the optimal policy is described by a single barrier above which dividends are paid. The value of the firm then reduces to one of a firm in the absence of the capital issue option.

The assumption of constant scale in a firm's operational cash flow could seem overly restrictive. We present two arguments in favor of the view that constant scale is a very relevant benchmark case. First, we present in the Appendix a model with stationary growth which reduces to our model with constant scale through a simple normalization. Second, we note here that there are theoretical results which assure that there is some robustness in the assumption of constant scale. In particular, Radner and Shepp (1996) have shown that when constant scale (in this type of model but without capital issues) is replaced with constant returns to scale, the basic problem no longer leads to a reasonable solution, but to one where either all capital is paid as dividends immediately (the drift is less than the discount rate), or dividends are withheld indefinitely and the value of the objective is infinite (the drift is higher than the discount rate). This reflects the mathematical incompatibility of linear discounting and constant returns to scale. Moreover, Alvarez (2001) and Radner (1998) have analyzed a variation of the basic model with linear utility where technology experiences positive scale effects that ultimately converge to a sufficiently low value, and has found that the optimal dividend policy is the same type of barrier policy than in the presence of constant scale. Constant scale is therefore not a necessary condition for the qualitative results of the model,

and qualitatively similar results are obtained from any model with e.g. declining and ultimately vanishing returns to scale.

There is a sizable literature on the singular stochastic control of (arithmetic) Brownian motion. The references that are closest to our work have already been mentioned in this Introduction. Other areas where singular control has been applied include inventory control (Harrison and Taksar, 1983, Harrison, 1985), portfolio optimization under proportional transaction costs (Constantinides, 1986, Davis and Norman, 1989), and optimal firing and hiring (Bentolila and Bertola, 1990, Shepp and Shiryaev, 1996).

The rest of the paper is organized as follows. Section 2 presents our model. Optimal policies and the value function are derived in section 3. Numerical illustrations and comparative static analyses are in section 4. Section 5 concludes with a number of remarks concerning possible extensions and variations of the model presented here.

2 The model

We analyze a firm with a cumulative profit flow Y which evolves according to

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

where $\{W_t : t \geq 0\}$ is a standard Wiener process, and the parameters μ and σ are positive constants. The standard filtration generated by the Brownian motion $\{W_t : t \geq 0\}$ is denoted $\{F_t : t \geq 0\}$. This implies that instantaneous profit flow is non-predictable, and that both positive and negative instantaneous profits occur with positive probability.

A dividend-capital raising policy is a two-dimensional stochastic process (D, S) , where D_t is the cumulative amount of dividends paid up to time t and S_t is the cumulative amount of new capital raised up to time t . We denote by Π the set of policies (D, S) which are non-decreasing right-continuous processes, adapted to F_t , satisfy $D_{0-} = S_{0-} = 0$, and where S_t satisfies the constraint

$$S_t = \int_0^t s_u du, \quad 0 \leq s_t \leq \delta \text{ for all } t.$$

The firm's capital stock, as a function of a policy (D, S) , is denoted $X_t^{D,S}$ and evolves according to

$$\begin{aligned} dX_t^{D,S} &= dY_t - dD_t + (1-\alpha)dS_t \\ &= (\mu + (1-\alpha)s_t)dt + \sigma dW_t - dD_t. \end{aligned} \quad (2)$$

Any profits and losses feed to the capital stock, dividends are paid out of the capital stock, and issues of new equity add to the capital stock. The parameter $\alpha \in (0,1)$ is the proportional cost rate of external capital. We interpret this so that for each unit of new capital raised, α units are used to pay external parties which facilitate the collection of external capital, and $1 - \alpha$ units are retained to accumulate the firm's capital stock. The initial stock of capital X_0 is assumed to be non-negative and is generically denoted by x .

The firm operates until the capital stock for the first time hits zero, i.e. until the stopping time

$$\tau_{D,S} = \inf\{t \geq 0 : X_t^{D,S} \leq 0\},$$

where X evolves according to (2). Zero is assumed to be the (normalized) default boundary, the violation of which destroys the firm's potential to continue operating as a going concern.

Firm value under policy (D, S) , given an initial capital stock x , is the expected discounted present value of dividends less new capital issues until default time

$$V_{D,S}(x) = E \int_{0-}^{\tau_{D,S}} e^{-\rho t} (dD_t - dS_t). \quad (3)$$

where ρ is a positive parameter denoting the excess required return on equity over the riskless rate. The value function of the problem is defined by

$$V(x) = \sup_{D,S \in \Pi} V_{D,S}(x), \quad (4)$$

and corresponds to the value of an optimally managed firm. The problem is to identify V as defined in (4) and an optimal policy (D^*, S^*) such that $V(x) = V_{D^*, S^*}(x)$. We note that the problem has five parameters $(\mu, \sigma, \rho, \alpha, \delta)$. The first two of these describe the profit dynamics from the productive technology, the third is the risk premium on equity, and the last two describe the capital market imperfections which constrain the acquisition of external capital.

The model defined by (1-4) describes a firm whose capital dynamics does not depend on time or on the size of the firm's capital stock. Hence there are no returns to scale in our model. This firm may be interpreted as a bank which has access to an infinitely elastic supply of zero cost (insured) deposits, and which has an illiquid portfolio of a fixed size. This interpretation has been suggested by Milne and Whalley (2001) in the context of the same basic capital dynamics. The capital dynamics in (2), however, can also be derived from a more general capital dynamics which is neither time nor state homogenous, through an appropriate normalization. We present this extension of our model in Appendix A, and show how (2) and (3) are obtained from the more general model.

3 Value function and optimal policies

The problem defined by (4) is a mixed regular and singular control problem. That V is concave follows from the linearity of the objective and the linearity of the capital dynamics using standard arguments as illustrated e.g. in Hojgaard and Taksar (1999). Assuming that V is twice continuously differentiable on $(0, \infty)$, standard dynamic programming arguments (see e.g. Fleming and Soner (1993), or Hojgaard and Taksar (1999)) then imply that V satisfies the Hamilton-Jacobi-Bellman (HJB) variational equation

$$\max \left(\max_{0 \leq s \leq \delta} \left[\frac{1}{2} \sigma^2 V''(x) + (\mu + (1 - \alpha)s)V'(x) - \rho V(x) - s \right], 1 - V'(x) \right) = 0, \quad x > 0 \quad (5)$$

$$V(0) = 0. \quad (6)$$

Our main task in this section is to construct a twice continuously differentiable, concave solution to (5) and (6). We find it useful to separate the analysis of the general case where $\alpha \in$

$(0, 1)$ and $\delta \in (0, \infty)$ from the various limiting cases which are obtained as either $\alpha = 1$ or $\alpha = 0$, or as $\delta = 0$ or $\delta \rightarrow \infty$. These limiting cases are discussed in section 3.2.

3.1 The general case $\alpha \in (0, 1)$, $\delta \in (0, \infty)$

We seek to construct a concave solution to (5) that satisfies (6). We denote such a solution by f . Then f satisfies (5) written in an equivalent form

$$\max \left(\max_{0 \leq s \leq \delta} \left[\frac{1}{2} \sigma^2 f''(x) + \mu f'(x) - \rho f(x) + ((1-\alpha)f'(x) - 1)s \right], 1 - f'(x) \right) = 0, \quad (7)$$

$x > 0$. We define barriers b_1 and b_2 in terms of f as

$$b_1 = \inf \{x \geq 0 : (1-\alpha)f'(x) = 1\}, \quad (8)$$

$$b_2 = \inf \{x \geq 0 : f'(x) = 1\}. \quad (9)$$

By concavity and twice continuous differentiability of f , $b_1 < b_2$. We denote by $s(x)$ the function that achieves the inner maximum in (7) for any x . Then it follows from the concavity of f and (7) that

$$s(x) = \begin{cases} \delta, & x \leq b_1 \\ 0, & x > b_1 \end{cases}. \quad (10)$$

This implies that optimal policies have the following form: i) capital is issued at the maximal admissible rate δ when the capital stock is at or below b_1 , ii) no controls are applied when the capital stock is in the interval (b_1, b_2) , iii) dividends are paid at maximal admissible rate when the capital stock is above b_2 . Since dividends may be paid at an unbounded rate, the process for capital is reflected at b_2 .

The behavior of f can then be classified into three regions. For $0 < x < b_1$, f satisfies

$$\frac{1}{2} \sigma^2 f''(x) + (\mu + (1-\alpha)\delta)f'(x) - \rho f(x) - \delta = 0.$$

This has the general solution, denoted f_1 ,

$$f_1(x) = c_{11}e^{d_{1+}x} + c_{12}e^{d_{1-}x} - \frac{\delta}{\rho}, \quad (11)$$

where

$$d_{1\pm} = \frac{1}{\sigma^2} \left[-(\mu + (1-\alpha)\delta) \pm \sqrt{(\mu + (1-\alpha)\delta)^2 + 2\rho\sigma^2} \right]. \quad (12)$$

For $b_1 < x < b_2$, f satisfies

$$\frac{1}{2}\sigma^2 f''(x) + \mu f'(x) - \rho f(x) = 0,$$

which has a general solution, denoted f_2 ,

$$f_2(x) = c_{21}e^{d_{2+}x} + c_{22}e^{d_{2-}x}, \quad (13)$$

where

$$d_{2\pm} = \frac{1}{\sigma^2} \left[\mu \pm \sqrt{\mu^2 + 2\rho\sigma^2} \right]. \quad (14)$$

For $x > b_2$, f satisfies

$$f_3(x) = x + c_3. \quad (15)$$

We therefore conjecture the following solution to (7)

$$f(x) = \begin{cases} f_1(x) = c_{11}e^{d_{1+}x} + c_{12}e^{d_{1-}x} - \delta/\rho & 0 < x < b_1 \\ f_2(x) = c_{21}e^{d_{2+}x} + c_{22}e^{d_{2-}x} & b_1 < x < b_2 \\ f_3(x) = x + c_3 & x > b_2 \end{cases}$$

The five yet unknown constants (c_{11} , c_{12} , c_{21} , c_{22} , c_3) and the two endogenous boundaries (b_1 , b_2) are to be solved from the value matching and smooth pasting conditions associated with the problem (see Dumas (1991) or Dixit (1991) for a discussion of these). We note that the + and - signs in the known constants d_{1+} , d_{1-} , d_{2+} , d_{2-} indicate the signs of these constants. We will use this information repeatedly below when we derive the signs of the other constants.

The smooth pasting conditions at b_2 are

$$f_2'(b_2) = f_3'(b_2) = 1$$

$$f_2''(b_2) = f_3''(b_2) = 0.$$

Solving this system for c_{21} and c_{22} , and inserting these into (13), yields

$$f_2(x) = A_{21}e^{d_{2+}(x-b_2)} + A_{22}e^{d_{2-}(x-b_2)}, \quad (16)$$

$$A_{21} = \frac{d_{2-}}{d_{2+}(d_{2-} - d_{2+})}, \quad (17)$$

$$A_{22} = \frac{d_{2+}}{d_{2-}(d_{2+} - d_{2-})}. \quad (18)$$

The signs of the d 's indicate that $A_{21} > 0$ and $A_{22} < 0$. Then evaluating (16) at b_2 yields $f_2(b_2) = A_{21} + A_{22} = \mu/\rho$. A value matching condition at b_2 now determines c_3 , so that (15) becomes

$$f_3(x) = \frac{\mu}{\rho} + (x - b_2).$$

When b_1 is positive, f_2 satisfies the boundary condition $f_2'(b_1) = (1 - \alpha)^{-1}$ by definition (8) of b_1 . Substituting (16) into this condition yields an equation that implicitly determines $b_2 - b_1$

$$A_{21}d_{2+}e^{-d_{2+}(b_2-b_1)} + A_{22}d_{2-}e^{-d_{2-}(b_2-b_1)} = \frac{1}{1 - \alpha}. \quad (19)$$

We cannot solve this explicitly for $b_2 - b_1$, but the following lemma establishes the existence of a unique positive solution to (19).

Lemma 1. *There exists a unique positive $b_2 - b_1$ that solves (19).*

Proof: Define the continuous function $g: [0, \infty) \rightarrow \mathbb{R}$ by

$$g(y) = A_{21}d_{2+}e^{-d_{2+}y} + A_{22}d_{2-}e^{-d_{2-}y}. \quad (20)$$

We record the following properties of g : 1) $g(y) > 0 \forall y \geq 0$; 2) $g(0) = 1$; 3) $g'(0) = 0$; 4) $g''(y) > 0 \forall y \geq 0$; 5) g is unbounded.

1) holds since $A_{21} > 0$ and $A_{22} < 0$. To show 2) and 3), we evaluate g and g' at 0 and use the expressions for A_{21} and A_{22}

$$g(0) = A_{21}d_{2+} + A_{22}d_{2-} = -\frac{d_{2-}}{d_{2+} - d_{2-}} + \frac{d_{2+}}{d_{2+} - d_{2-}} = \frac{d_{2+} - d_{2-}}{d_{2+} - d_{2-}} = 1$$

$$g'(0) = -A_{21}d_{2+}^2 - A_{22}d_{2-}^2 = \frac{d_{2+}d_{2-}}{d_{2+} - d_{2-}} - \frac{d_{2+}d_{2-}}{d_{2+} - d_{2-}} = 0$$

To show 4), we differentiate g twice and determine the signs of the coefficients. 5) follows from the fact that the exponential function is unbounded, and $-d_{2-} > 0$. Properties 3 and 4 imply that g is increasing, so that for $\alpha > 0$, g crosses $(1-\alpha)^{-1}$ at a single point $y^*(\alpha) > 0$, which is the value of $b_2 - b_1$ that solves (20). End of proof.

We observe from (19) that $b_2 - b_1$ does not depend on δ , but is driven by the cost rate α . The next result characterizes the dependence of $b_2 - b_1$ on α .

Lemma 2. $b_2 - b_1 = 0$ when $\alpha = 0$, $b_2 - b_1$ is an increasing function of α , and

$$\left. \frac{\partial(b_2 - b_1)}{\partial \alpha} \right|_{\alpha=0} = \infty.$$

Proof: Let g be the function defined in (20). The first claim follows since $g(0) = 1$. By totally differentiating (19) w.r.t. $b_2 - b_1$ and α , we obtain

$$\frac{\partial(b_2 - b_1)}{\partial \alpha} = \frac{1}{(1-\alpha)^2} \frac{1}{g'(y^*)} > 0,$$

when $\alpha \in (0,1)$, since $g'(y) > 0$ when $y > 0$, and $y^*(\alpha) > 0$ when $\alpha > 0$. The third claim follows since $g'(0) = 0$ and $y^*(0) = 0$. End of proof.

As α approaches 1, the right-hand side of (19) approaches infinity. The value of $b_2 - b_1$ solving (19) therefore also increases without bound. This solution, however, is valid only for those values of α which are consistent with f_2 at the dividend barrier b_1 being positive. The following lemma gives an upper bound on α .

Lemma 3. $f_2(b_1) > 0$ when $\alpha < \hat{\alpha}$, where $\hat{\alpha}$ is given by

$$\hat{\alpha} = 1 - \left(-\frac{d_{2-}}{d_{2+}} \right)^{\left(\frac{d_{2+} + d_{2-}}{d_{2+} - d_{2-}} \right)}. \quad (21)$$

Proof. We know that $f_2(b_2) = \mu/\rho$ for any b_2 , and the left-hand side of (19) shows that the partial derivative of $f_2(b_1)$ with respect to $(b_2 - b_1)$ is negative. Therefore there is a critical value b_0 for $b_2 - b_1$ which yields $f_2(b_1) = 0$. Substituting (16) into this condition yields

$$A_{21}e^{-d_{2+}b_0} + A_{22}e^{-d_{2-}b_0} = 0, \quad (22)$$

from which b_0 can be solved explicitly. Since $(b_2 - b_1)$ is increasing in α , corresponding to b_0 there is a critical value of α , denoted $\hat{\alpha}$, such that b_0 solves (19) when α is at its critical value, i.e.

$$A_{21}d_{2+}e^{-d_{2+}b_0} + A_{22}d_{2-}e^{-d_{2-}b_0} = \frac{1}{1 - \hat{\alpha}}.$$

Substituting in (22) and solving for $\hat{\alpha}$ gives (21). End of proof.

We observe from (21) that $\hat{\alpha} \in (0,1)$ since $-d_{2-} > d_{2+} > 0$. Moreover, $\hat{\alpha}$ is a function of μ , σ and ρ , but does not depend on δ . This may appear a little surprising. One could a priori expect that higher capital issuance rates would cause the firm to tolerate higher costs of capital issuance.

When α equals its critical value, the marginal value of capital at 0 just equals the hurdle rate for issuance of costly external capital, $(1-\alpha)^{-1}$. Figure 1 graphs the critical value $\hat{\alpha}$ as a function of σ , for selected values of the drift rate μ . The figure indicates that the critical value of α is a declining function of the cash flow volatility, and an increasing function of the cash flow drift. As volatility converges to zero, the critical value converges to one. The critical value is also remarkably high at such values of the drift and the volatility parameters which may be deemed typical. When the (μ, σ, ρ) triple is at $(1, 2, 0.1)$, e.g., the critical value for α is 0.76, implying that the marginal value of capital at zero is 4.2.

We continue with the solution of the general case where $\alpha < \hat{\alpha}$. Given the solution (16) for f_2 as a function of b_2 , we solve c_{11} and c_{12} from the smooth pasting conditions at b_1 ,

$$f_1'(b_1) = f_2'(b_1),$$

$$f_1''(b_1) = f_2''(b_1).$$

Solving this pair of linear equations for c_{11} and c_{12} and inserting these into (11) yields

$$f_1(x) = A_{11}e^{d_{1+}(x-b_1)} + A_{12}e^{d_{1-}(x-b_1)} - \frac{\delta}{\rho}, \quad (23)$$

$$A_{11} = A_{21} \frac{d_{2+}(d_{2+} - d_{1-})}{d_{1+}(d_{1+} - d_{1-})} e^{-d_{2+}(b_2-b_1)} + A_{22} \frac{d_{2-}(d_{2-} - d_{1-})}{d_{1+}(d_{1+} - d_{1-})} e^{-d_{2-}(b_2-b_1)}, \quad (24)$$

$$A_{12} = A_{21} \frac{d_{2+}(d_{2+} - d_{1+})}{d_{1-}(d_{1-} - d_{1+})} e^{-d_{2+}(b_2-b_1)} + A_{22} \frac{d_{2-}(d_{2-} - d_{1+})}{d_{1-}(d_{1-} - d_{1+})} e^{-d_{2-}(b_2-b_1)}. \quad (25)$$

Analysis of the signs of the coefficients shows that $A_{11} > 0$ and $A_{12} < 0$.

In (24) and (25), $b_2 - b_1$ is assumed to be the unique positive solution to (19). b_1 is therefore the only unknown, and can be determined from the remaining boundary condition $f_1(0) = 0$. Substituting (23) into this condition yields an equation that implicitly determines b_1 ,

$$A_{11}e^{-d_{1+}b_1} + A_{12}e^{-d_{1-}b_1} = \frac{\delta}{\rho}. \quad (26)$$

The following lemma verifies the existence of a unique solution to (26).

Lemma 4. *A unique solution b_1 to (26) exists. The solution is strictly positive when $\alpha < \hat{\alpha}$.*

Proof: We define the function $h: R \times (0,1) \rightarrow R$ by $h(y, \alpha) = A_{11}(\alpha)e^{-d_{1+}y} + A_{12}(\alpha)e^{-d_{1-}y}$, where A_{11} and A_{12} have been defined by (24) and (25), and do not depend on y . The argument emphasizes their dependence on α .

I Existence and uniqueness. We know that $A_{11}(\alpha) > 0$ and $A_{12}(\alpha) < 0$ for all $\alpha \in (0, 1)$. Differentiation of h w.r.t. y then shows that $h_1(y, \alpha) < 0$ for all y . Moreover, $h(y, \alpha) \rightarrow \infty$ as $y \rightarrow -\infty$, and $h(y, \alpha) \rightarrow -\infty$ as $y \rightarrow \infty$. This implies that for all $\alpha \in (0, 1)$, there exists a unique $y^*(\alpha)$ such that $h(y^*(\alpha), \alpha) = \delta/\rho$.

II Positivity. We note that $f_2(b_1) > 0$ when $\alpha < \hat{\alpha}$ by Lemma 3. Because of value matching at b_1 , $f_1(b_1) = f_2(b_1) > 0$. Value matching at b_1 follows from the smooth pasting conditions at

b_1 . Differentiation of (23), on the other hand, shows that $f_1'(x) > 0$ for all x . This, combined with $f_1(0) = 0$ (condition (26)) and $f_1(b_1) > 0$ implies that b_1 is positive. End of proof.

We have now determined the constants $(c_{11}, c_{12}, c_{21}, c_{22}, c_3)$ as explicit functions of the endogenous barriers (b_1, b_2) , and the two endogenous barriers (b_1, b_2) as implicit functions of the problem parameters. To summarize, when $0 < \alpha < \hat{\alpha}$, we propose the following solution to (7) subject to $f(0) = 0$,

$$f(x) = \begin{cases} A_{11}e^{d_{1+}(x-b_1)} + A_{12}e^{d_{1-}(x-b_1)} - \frac{\delta}{\rho} & 0 \leq x \leq b_1 \\ A_{21}e^{d_{2+}(x-b_2)} + A_{22}e^{d_{2-}(x-b_2)} & b_1 < x < b_2, \\ \frac{\mu}{\rho} + (x - b_2) & x \geq b_2 \end{cases} \quad (27)$$

where b_1 solves (26), $b_2 - b_1$ solves (19), A_{11} and A_{12} are given by (24) and (25), and A_{21} and A_{22} are given by (17) and (18). The next lemma verifies that (27) is a concave solution to (7).

Lemma 5. *f given by (27) is a concave, twice continuously differentiable solution to (7) subject to $f(0) = 0$.*

Proof: We first show that f defined by (27) is concave. We know that $f'(x) = 0$ for all $x \geq b_2$, and f is therefore (weakly) concave on $[b_2, \infty)$. The smooth pasting conditions at b_2 imply that $f_2''(b_2) = 0$. Differentiating f_2 given by (16) three times, and analyzing the signs of the coefficients, shows that $f_2'''(x) > 0$ on (b_1, b_2) . f therefore has an increasing second derivative on (b_1, b_2) , which, combined with the fact that $f'(b_2) = 0$, implies that $f'(x) < 0$ on (b_1, b_2) . Thus f is concave on (b_1, b_2) . Similar reasoning can be used to establish the concavity of f on $(0, b_1)$. Differentiating f_1 given by (23) three times and using the established fact that $A_{11} > 0$, $A_{12} < 0$, shows that $f_1'''(x) > 0$ on $(0, b_1)$. Again, this combined with the fact that $f'(b_1) < 0$ because of smooth pasting with f_2 , implies that $f'(x) < 0$ on $(0, b_1)$.

(27) is twice continuously differentiable by construction. By concavity, $f(x) > 1$ for $x < b_2$, so that $1 - f(x) < 0$ and (27) satisfies (7) for $x < b_2$ by construction. Since $f(x) = 1$ for $x \geq b_2$, we need to show that $L^s f(x) \leq 0$ for $x \geq b_2$, for all $s \in [0, \delta]$, where

$$L^s f(x) = \frac{1}{2} \sigma^2 f''(x) + \mu f'(x) - \rho f(x) + ((1 - \alpha)f'(x) - 1)s.$$

Substituting (27) into this, we get

$$\begin{aligned}
L^s f(x) &= \mu - \rho \left(\frac{\mu}{\rho} + (x - b_2) \right) - \alpha s \\
&= -\rho(x - b_2) - \alpha s \leq 0
\end{aligned}$$

for all $s \in [0, \delta]$. End of proof.

By a sufficiency argument similar to the one proved in Hojgaard and Taksar (1999), the solution (27) coincides with the value function defined in (4). Moreover, an optimal barrier policy exists. We state the result in the following proposition, but will not show the proof since the argumentation is essentially identical to the proof in Hojgaard and Taksar (1999).

Proposition 1. *Let $0 < \alpha < \hat{\alpha}$, and $f(x)$ be defined by (27). Then $f(x) = V(x) = V_{D^*, S^*}(x)$ for all x , where (D^*, S^*) is the policy defined by the system of equations*

$$X_t^{D^*, S^*} = x + \mu t + (1 - \alpha) \int_0^t s(X_u^{D^*, S^*}) du + \sigma W_t - D_t^*,$$

$$X_t^{D^*, S^*} \leq b_2, t \geq 0,$$

$$\int_0^\infty \mathbf{1}(X_t^{D^*, S^*} < b_2) dD_t^* = 0,$$

where $s(x)$ is the function defined in (10), and b_1 and $b_2 - b_1$ are the solutions to (27) and (19).

3.2 The limiting cases

We can distinguish three different limiting cases to our model, according to the nature of the resulting solution, as the two parameters α and δ describing capital market imperfections approach their limiting values.

Case I: $\delta = 0$ or $\alpha \geq \hat{\alpha}$

In the limiting case $\delta = 0$ the capital issue option is not available. Also, it was shown in the previous section that whenever α is equal or greater to the critical value given by (21), the optimal policy does not involve capital issues. In both cases the solution reduces to the special case which has been solved by Milne and Robertson (1996), among others. The optimal policy is now described in terms of a single barrier b_0 , which is a reflecting dividend barrier

for the capital stock. The barrier b_0 satisfies the condition $f_2(0)=0$, which is written out in (22), and can be solved for b_0 explicitly as

$$b_0 = \frac{2}{d_{2+} - d_{2-}} \ln\left(-\frac{d_{2-}}{d_{2+}}\right), \quad (28)$$

where d_{2+} and d_{2-} are as in (14). The resulting value function is

$$V(x) = \begin{cases} A_{21}e^{d_{2+}(x-b_0)} + A_{22}e^{d_{2-}(x-b_0)} & 0 \leq x < b_0 \\ \frac{\mu}{\rho} + (x - b_0) & x \geq b_0 \end{cases}. \quad (29)$$

Case II: $\alpha = 0$

As $\alpha \rightarrow 0$, it follows from the definitions (8) and (9) and continuous differentiability of f that $b_2 - b_1 \rightarrow 0$. Lemma 2 in fact shows that $b_2 - b_1 = 0$ when $\alpha = 0$, so that the optimal policy in this case is characterized by a single barrier b^* , above which dividends are paid at unlimited rate (so that the process is reflected at b^*), and below which new capital is raised at the maximal admissible rate δ . The value function and the barrier b^* are characterized by the following result, whose derivation proceeds as in the general case.

Proposition 2. *Let $\alpha = 0$. The dividend barrier b^* is the unique positive solution to the equation*

$$A_1 e^{-d_+ b^*} + A_2 e^{-d_- b^*} = \frac{\delta}{\rho}, \quad (30)$$

and the value function is

$$V(x) = \begin{cases} A_1 e^{d_+(x-b^*)} + A_2 e^{d_-(x-b^*)} - \frac{\delta}{\rho} & 0 \leq x < b^* \\ \frac{\mu}{\rho} + (x - b^*) & x \geq b^* \end{cases} \quad (31)$$

where

$$A_1 = \frac{d_-}{d_+(d_- - d_+)} > 0, \quad A_2 = \frac{d_+}{d_-(d_+ - d_-)} < 0,$$

$$d_{\pm} = \frac{1}{\sigma^2} \left[-(\mu + \delta) \pm \sqrt{(\mu + \delta)^2 + 2\rho\sigma^2} \right].$$

Case III: $\delta = \infty$

As δ increases, the firm's control on the minimum level of the capital stock improves. In the limit, as δ is unbounded, the control on the minimum level of the capital stock is of 'barrier' type, so that the capital stock is reflected at b_1 . In this situation the owners would clearly prefer to set the barrier b_1 as close to zero as possible. Yet the barrier cannot equal zero, since this would not be enough to prevent bankruptcy. Unsurprisingly, then, optimal policy does not exist in the case where the rate of capital injection is unbounded.

We can derive the 'unreachable' value function by the following heuristic argument. We know that when the rate of capital injections is unbounded, the capital injection barrier b_1 is arbitrarily close to zero. Then a decision to raise a small amount Δ of new capital at b_1 raises the capital stock to $b_1 + (1 - \alpha)\Delta$. Since this is the approximately optimal action at barrier b_1 , the value function at b_1 equals the net value from this operation. Thus we have

$$\begin{aligned} V(b_1) &\approx V(b_1 + (1 - \alpha)\Delta) - \Delta \\ &\approx V(b_1) + V'(b_1)(1 - \alpha)\Delta - \Delta \end{aligned}$$

Solving the previous approximate equality for $V(b_1)$, and letting both b_1 and Δ approach zero yields

$$V'(0) = \frac{1}{1 - \alpha}. \quad (32)$$

The value function in the case $\delta = \infty$ can be found by solving the HJB equation (5) subject to the boundary condition (32). The solution, which can be obtained by straightforward adaptation of earlier arguments, is given in the following proposition.

Proposition 3. *Let $\delta = \infty$. When $\alpha < \hat{\alpha}$, the value function is given by*

$$V(x) = \begin{cases} A_{21}e^{d_{2+}(x-\hat{b})} + A_{22}e^{d_{2-}(x-\hat{b})} & 0 \leq x < \hat{b} \\ \frac{\mu}{\rho} + (x - \hat{b}) & x \geq \hat{b} \end{cases} \quad (33)$$

where A_{21} and A_{22} are given by (17) and (18), and the dividend barrier \hat{b} is the unique positive solution to (19). Else, the solution is given by (28) and (29).

The condition on α in the proposition means that even though δ approaches infinity, capital issues may not be optimally used when the cost rate is sufficiently high. The critical value of α is in fact the same as in the general model with finite δ . This is consistent with our earlier finding that $\hat{\alpha}$ does not depend on δ . When α is below the critical value, the barrier \hat{b} solves the same equation as the difference $b_2 - b_1$ in the general model. Also the value function (33) is simply a left-shifted version of the value function (27) of the general model with finite δ , to the right of b_1 . In other words, as δ approaches infinity, b_1 approaches zero, but the difference $b_2 - b_1$ remains constant. We noted already in connection with Lemma 2 that $b_2 - b_1$ does not depend on δ .

We finally look at what happens when α approaches zero and δ simultaneously approaches infinity, i.e. both capital market frictions disappear. This limit corresponds to frictionless capital markets, and as such is an interesting benchmark. Based on our previous results, in this case $b_2 - b_1$ approaches zero (due to vanishing α) and b_1 approaches zero (due to δ approaching infinity). Both b_1 and b_2 therefore converge to zero, and the optimal policy converges to one where the capital stock is trapped at zero indefinitely. Of course, the optimal policy in the limiting case does not exist, but ε -optimal policies which keep the capital stock (most of the time) between two barriers, set arbitrarily close to each other as well as arbitrarily close to zero, can be implemented. The value function in the limiting case is obtained as the limit of (31) when b^* approaches zero, or alternatively, as a limit of (33) when \hat{b} approaches zero, and takes the affine form

$$V(x) = \frac{\mu}{\rho} + x.$$

4 Numerical illustrations

4.1 The value of the capital issue option

The value function (27) is the value of an optimally managed firm which has access to costly external capital. The value function (29) is the value of an otherwise identical firm which does not have access to external capital. The difference between (27) and (29) is the value of the option to issue new capital.

Figure 2 illustrates both value functions for different values of the parameter α . The cost of capital injections is zero in picture (1), 0.2 in picture (2), 0.4 in picture (3), and 0.6 in picture (4). The maximal rate of capital injections in all pictures is equal to the volatility rate. The value of the option to issue new capital declines with the cost rate α . This is because the value (27) of the firm with the capital issue option declines with the cost rate, while the value of the non-optional firm does not depend on the cost rate.

Figure 3 shows the value of the capital issue option for the cases considered in Figure 2. The upper picture shows the absolute value of the capital issue option, while the lower picture shows the value of the option as a percentage of the cum-option firm value. Concerning the absolute value of the capital issue option, we observe a pattern that appears to be qualitatively robust to variations in the parameter values. First, the capital issue option has little value when the capital stock is very close to zero. The volatility of a diffusion locally dominates its drift, and therefore capital issues which are just changes in the drift of the diffusion process for the capital stock, have little effect on near-term bankruptcy probability. At a reasonable distance from zero, however, changes in the drift rate begin to have a more significant effect on the near-term probability of bankruptcy, and the value of the capital issue option therefore increases with the initial capital level. The higher the capital stock, on the other hand, the less likely will be near-term bankruptcy, even in the absence of the additional drift due to capital issuance. The value of the capital issue option therefore peaks at an intermediate level of the capital stock, and then begins to decline, ultimately converging to a non-negative constant. Specifically, the value of the capital issue option is constant and equal to $b_0 - b_2$ when the capital stock is above b_0 , which is obtained through subtracting (29) from (27). This value is non-negative since the barrier b_2 in the general model may never exceed the barrier b_0 in the model without capital issues. This result highlights the fact that the value of the capital issue option derives from the ability to advance dividend payments, i.e. to reduce the barrier b_2 relative to the barrier b_0 .

The lower part of Figure 3 shows that the value of the capital issue option, as a proportion of the cum-option firm value, declines monotonically with respect to the capital stock. We see that the option to issue new capital can more than double the firm value when the capital stock is low and the cost rate α is not prohibitively high. As the capital stock increases without bound, on the other hand, the proportional value of the option converges to zero. This is because the absolute value of the option converges to a finite constant (as discussed above)

while the value of the firm is unbounded. The proportional value of the option when the capital stock is at the dividend barrier b_2 measures the value of the capital issue option for an optimally capitalized firm. At this point the value of the optional firm is μ/ρ , while the value of the non-optional firm is obtained by evaluating (29) at b_2 . In the cases $\alpha = 0.0$ and $\alpha = 0.2$ in Figure 3, the corresponding value of the option is 16% and 7%, respectively. These percentages represent the portion of firm value attributable to the flexibility to issue new capital, for a firm which is optimally capitalized (capital stock equals the dividend barrier).

4.2 Sensitivity of b_1 and b_2 to capital market imperfections

We have shown in the previous section that the optimal policy is to issue capital at the maximal admissible rate when the capital stock is below b_1 , and to pay dividends so as to prevent capital stock from rising above b_2 . The barriers b_1 and b_2 depend on the magnitude of the capital market imperfections described by the parameters α and δ . Here we illustrate this dependence.

Figure 4 shows b_1 and b_2 as a function of the cost rate α . For a given δ , $b_2 - b_1$ is an increasing function of α (Lemma 2). Moreover, $b_2 - b_1$ does not depend on δ , so that changing δ either increases or decreases b_1 and b_2 by equal amounts, given a fixed α . When α approaches its critical value defined in (21) from below, b_1 approaches 0 and b_2 approaches b_0 . It is noteworthy that $\hat{\alpha}$ is the smallest value of α at which b_1 equals 0, and also the smallest value of α at which b_2 equals b_0 .

The most surprising aspect of Figure 4 is the fact that b_1 may sometimes increase with α . In particular, when δ is relatively high, such as the case where $\delta = 30$, there is a region of α -values over which b_1 is increasing. There are two counteracting effects which play a role here. On one hand, a larger α makes capital issues more expensive, suggesting that the capital issue barrier b_1 should be lowered. On the other hand, a larger α means that the net rate of capital issuance $\delta(1-\alpha)$ is reduced, suggesting an increase in b_1 . It turns out that for relatively low values of δ , the former effect dominates, while for high δ the latter effect may dominate, which is what we observe in Figure 4.

The capital issue barrier b_1 is non-increasing with respect to δ in all the cases considered in Figure 4, although we have not proven this in general. A higher δ allows the firm to wait

longer before initiating capital issues. As δ converges to infinity, the capital issuance barrier b_1 converges to zero, and the dividend barrier b_2 converges to the positive value solving (19). This limiting case has been analyzed in Proposition 3.

4.3 Comparison of capital issues with temporary risk reductions

In this section we compare the value of capital issues with the value of temporary risk reductions, which have been analyzed by Hojgaard and Taksar (HT) (1999). Their basic model is the same as in this paper, but instead of capital issues, HT allow the firm to reduce the volatility of its cash flow, in which case the drift of the cash flow is reduced proportionally. The controlled capital stock in their model evolves according to

$$dX_t^{a,D} = \mu a_t dt + \sigma a_t dW_t - dD_t,$$

where the coefficient of risk reduction a is allowed to take values in the interval $[0, 1]$. The optimal risk reduction policy in the HT model is the following: for $x \in [0, u_1]$,

$$a(x) = -\frac{\mu x}{\sigma^2(\gamma - 1)}$$

and for $x > u_1$, $a(x) = 1$, where

$$u_1 = \frac{\sigma^2}{\mu}(1 - \gamma),$$

$$\gamma = \frac{\rho}{\mu^2/2\sigma^2 + \rho}.$$

The dividend policy is the same form of barrier policy as in our model, so that every excess above boundary u_2 is distributed away instantly, where the boundary u_2 is given by

$$u_2 = u_1 + \frac{1}{d_{2+} - d_{2-}} \ln\left(\frac{-d_{2-}}{d_{2+}}\right). \quad (34)$$

The value function in the HT model, using our notation, is

$$V(x) = \begin{cases} \frac{2\lambda\mu}{\sigma^2} \left(\frac{x}{u_1}\right)^\gamma & 0 \leq x \leq u_1 \\ A_{21}e^{d_{2+}(x-u_2)} + A_{22}e^{d_{2-}(x-u_2)} & u_1 < x < u_2, \\ \frac{\mu}{\rho} + (x - u_2) & x \geq u_2 \end{cases} \quad (35)$$

where A_{21} and A_{22} are as defined in (17) and (18).

Figure 5 plots three value functions: i) the value function of the basic model in the absence of risk reductions and capital injections (dotted line), given in (29); ii) the HT value function in the presence of risk reductions (broken line), given in (35); iii) the value function with capital injections (solid line), given in (27). The option to issue new capital or to reduce risk cannot reduce firm value, so that the value functions associated with these alternatives dominate the value function of the basic model. Moreover, we observe from Figure 5 that in general, neither of the two risk management techniques is superior to each other in terms of their effects on firm value. When the maximal rate of capital issuance is low, such as in pictures (1) and (3), the value of risk reductions exceeds that of capital issues, irrespective of the level of the capital stock. This outcome changes, however, when the maximal rate of capital issuance increases. In pictures (2) and (4) new capital may be obtained at a rate that is three times the asset volatility rate. Here the value of risk reductions dominates the value of capital issues at low capital levels, but at higher levels of the capital stock the situation is reversed. It is intuitively clear that risk reductions are relatively most valuable at very low levels of the capital stock. This is because the volatility of a diffusion locally dominates its drift, and bankruptcy can only be avoided if the diffusion can be turned off. Capital issues only control for the drift of the diffusion, and therefore cannot completely eliminate the possibility of bankruptcy, given bounded drift rates. Risk reductions, on the contrary, under certain conditions are able to completely eliminate bankruptcies (bankruptcy under the optimally controlled cash flow process then becomes a zero probability event).

The probability of bankruptcy under optimal policies can be analyzed with the help of the scale function associated with the controlled diffusion. Karlin and Taylor (1981) show that the probability of bankruptcy in finite time is zero if the scale function of the diffusion, which in the context of our model is defined by

$$\int_0^x e^{-2 \int_0^z \frac{\mu(y)}{\sigma^2(y)} dy} dz,$$

takes an infinite value for some $x > 0$. Hojgaard and Taksar (1999), based on this result, show that in their model the probability of bankruptcy in finite time is zero whenever

$$\frac{\mu^2}{2\sigma^2} > \rho.$$

It is easy to show that in the model of this paper the scale function is finite for all positive x . The controlled capital stock is reflected at a finite barrier b_2 , has a bounded drift, and a volatility coefficient bounded from below. It is well known (and can be shown using the methods in Karlin and Taylor, 1981) that under these conditions the probability of bankruptcy in finite time equals 1.

We finally look at how the dividend barrier in the model with capital issues is related to the dividend barrier in the model with risk reductions. Figure 6 provides a comparison of the dividend barriers in the three models. The dividend barrier b_0 in the basic model is an upper bound on the dividend barriers in the models with capital issues and risk reductions. We observe that the dividend barrier with risk reductions, u_2 , is approximately 10% lower than b_0 . These barriers do not depend on α or δ , while the dividend barrier with capital issues b_2 is a function of both α and δ . As δ increases and α is below its critical value, b_2 falls. When α is sufficiently low, b_2 ultimately (for sufficiently high δ) falls below the barrier u_2 . The lower is α , the lower is the value of δ at which b_2 intersects u_2 . When α is close to the critical value $\hat{\alpha}$, however, b_2 does not fall below u_2 even though δ approaches infinity. This is because $b_2 - b_1$ does not depend on δ and will exceed u_2 when α is sufficiently close to $\hat{\alpha}$.

5 Extensions

5.1 Direct bankruptcy costs

The model in this paper embeds an indirect or opportunity cost of bankruptcy. This is the value lost due to irreversible discontinuation of operations which are fundamentally profitable on average. We can add direct bankruptcy costs to the model with little added difficulty. Let us assume that a positive cost K is incurred upon bankruptcy. The problem structure then

remains unchanged, with the exception that the boundary condition upon bankruptcy, $V(0) = 0$, is replaced by $V(0) = -K$. This affects b_1 through condition (26), but the condition (19) determining $b_2 - b_1$ remains unchanged. It is easy to show that b_1 is non-decreasing with respect to K , which is in line with immediate intuition.

5.2 Different capital market imperfections

The proportional cost and the bounded rate of capital issuance may be replaced with alternative capital market imperfections. One alternative set of assumptions, consisting of a fixed cost of capital issuance and a delay in implementing the capital issue, is analyzed in Peura and Keppo (2002).

Appendix A

We present here a model with stationary growth in time, and show how this model, through a simple normalization, reduces to the model presented in Section 2. In the extended model, the scale of the cumulative profit flow grows at a constant (risk free) rate in time,

$$d\hat{Y}_t = \mu e^{rt} dt + \sigma e^{rt} dW_t, \quad (\text{A1})$$

where $\{W_t : t \geq 0\}$ is a standard Wiener process. The parameters μ and σ and the riskless rate r are positive constants.

As in the model of Section 2, a dividend-capital raising policy is a two-dimensional stochastic process (\hat{D}, \hat{S}) , where \hat{D}_t is the cumulative amount of dividends paid up to time t , and \hat{S}_t is the cumulative amount of new capital raised up to time t . We denote by $\hat{\Pi}$ the set of policies (\hat{D}, \hat{S}) which are non-decreasing right-continuous processes, adapted to F_t , satisfy $\hat{D}_{0-} = \hat{S}_{0-} = 0$, and where \hat{S}_t satisfies

$$\hat{S}_t = \int_0^t \hat{s}_u du, \quad 0 \leq \hat{s}_t \leq \delta e^{rt} \text{ for all } t.$$

This implies that the upper bound on the capital issue rate also grows at the risk free rate. The firm's capital stock as a function of policy (\hat{D}, \hat{S}) is denoted $\hat{X}_t^{\hat{D}, \hat{S}}$ and evolves according to

$$d\hat{X}_t^{\hat{D}, \hat{S}} = r\hat{X}_{t-}^{\hat{D}, \hat{S}} dt + dY_t - d\hat{D}_t + (1 - \alpha)d\hat{S}_t. \quad (\text{A2})$$

The capital stock itself in this model earns the riskfree rate. The dynamics for the capital stock in (A2) is neither time nor state homogenous. The initial stock of capital \hat{X}_0 is assumed to be non-negative and is generically denoted by x .

The firm operates until the capital stock for the first time hits zero, i.e. until the stopping time

$$\hat{\tau}_{\hat{D}, \hat{S}} = \inf \{t \geq 0 : \hat{X}_t^{\hat{D}, \hat{S}} \leq 0\},$$

where X evolves according to (2).

The firm value under policy (\hat{D}, \hat{S}) , given an initial capital stock x , is the expected discounted present value of dividends less new capital issues until default time

$$\hat{V}_{\hat{D}, \hat{S}}(x) = E \int_{0-}^{\hat{\tau}_{\hat{D}, \hat{S}}} e^{-(\rho+r)t} (d\hat{D}_t - d\hat{S}_t). \quad (\text{A3})$$

where ρ is a positive excess required return on equity, over the risk free rate (we note that here the discount rate is different from that in the model of Section 2). The value function of the problem is defined by

$$\hat{V}(x) = \sup_{\hat{D}, \hat{S} \in \hat{\Pi}} \hat{V}_{\hat{D}, \hat{S}}(x), \quad (\text{A4})$$

and corresponds to the value of an optimally managed firm.

We now show how the model defined by (A1-A4) can be reduced to (1-4). In the context of the model (A1-A4), we define the normalized capital stock X_t as

$$X_t = \hat{X}_t e^{-rt}. \quad (\text{A5})$$

Also for each policy (\hat{D}, \hat{S}) in $\hat{\Pi}$, we define the corresponding normalized policy (D, S) by

$$D_t = \int_0^t e^{-ru} d\hat{D}_u, \quad S_t = \int_0^t e^{-ru} d\hat{S}_u, \quad \text{for all } t \geq 0. \quad (\text{A6})$$

We observe from (A6) and the definitions of $\hat{\Pi}$ and Π that if $(\hat{D}, \hat{S}) \in \hat{\Pi}$ then $(D, S) \in \Pi$, where (D, S) has been obtained from (\hat{D}, \hat{S}) through (A6).

Now applying Ito's formula to (A5), using (A1), (A2) and (A6), yields

$$\begin{aligned} dX_t^{\hat{D}, \hat{S}} &= \mu dt + \sigma dW_t - e^{-rt} d\hat{D}_t + (1-\alpha)e^{-rt} d\hat{S}_t \\ &= \mu dt + \sigma dW_t - dD_t + (1-\alpha)dS_t = dX_t^{D, S}, \end{aligned} \quad (\text{A7})$$

where (D, S) has been obtained from (\hat{D}, \hat{S}) through (A6). We also have that $\hat{X}_0 = X_0 = x$. Given (A6), the default times satisfy

$$\hat{\tau}_{\hat{D}, \hat{S}} = \inf\{t \geq 0 : \hat{X}_t^{\hat{D}, \hat{S}} \leq 0\} = \inf\{t \geq 0 : X_t^{D, S} \leq 0\} = \tau_{D, S}.$$

Therefore for any $(\hat{D}, \hat{S}) \in \hat{\Pi}$, we can rewrite (A3) as

$$\begin{aligned} \hat{V}_{\hat{D}, \hat{S}}(x) &= E \int_{0-}^{\hat{\tau}_{\hat{D}, \hat{S}}} e^{-(\rho+r)t} (d\hat{D}_t - d\hat{S}_t) = E \int_{0-}^{\hat{\tau}_{\hat{D}, \hat{S}}} e^{-\rho t} (e^{-rt} d\hat{D}_t - e^{-rt} d\hat{S}_t) \\ &= E \int_{0-}^{\tau_{D, S}} e^{-\rho t} (dD_t - dS_t) = V_{D, S}(x) \end{aligned} \quad (\text{A8})$$

which shows that the value of the objective (A3) in the extended problem, equals the value of the objective (3) in the original problem, once the policy (\hat{D}, \hat{S}) has been transformed through (A6) into a corresponding normalized policy. Now because $(\hat{D}, \hat{S}) \in \hat{\Pi}$ implies $(D, S) \in \Pi$, we have from (A8) that

$$\hat{V}(x) = \sup_{\hat{D}, \hat{S} \in \hat{\Pi}} V_{\hat{D}, \hat{S}}(x) \leq \sup_{D, S \in \Pi} V_{D, S}(x) = V(x). \quad (\text{A9})$$

Moreover, we can reverse the reasoning in (A6) and for each normalized policy (D, S) in Π define the corresponding policy $(\hat{D}, \hat{S}) \in \hat{\Pi}$ by

$$\hat{D}_t = \int_0^t e^{ru} dD_u, \quad \hat{S}_t = \int_0^t e^{ru} dS_u, \quad \text{for all } t \geq 0. \quad (\text{A10})$$

Then performing the analogous but reverse steps to (A7) and (A8) gives a reverse inequality to (A9). This makes it clear that $V(x) = \hat{V}(x)$, where $V(x)$ is defined by (4) and $\hat{V}(x)$ is defined by (A4). Therefore solving of the time-homogenous constant scale problem (4) also yields a solution to the time-dependent problem with constant returns to capital, (A4).

References

- Alvarez, L.H.R., 2001, Singular stochastic control, linear diffusions, and optimal stopping: a class of solvable problems, *SIAM Journal of Control and Optimization*, 39, 6, 1697-1710.
- Asmussen, S., M. Taksar, 1997, Controlled diffusion models for optimal dividend pay-out, *Insurance: Mathematics and Economics*, 20, 1-15.
- Asmussen, S., B. Hojgaard, M. Taksar, 2000, Optimal risk control and dividend distribution policies: example of excess-of-loss reinsurance for an insurance company, *Finance and Stochastics*, 4, 299-324.
- Bentolila, S., G. Bertola, 1990, Firing costs and labour demand: how bad is eurosclerosis?, *Review of Economic Studies*, 57, 381-402.
- Constantinides, G.M., 1986, Capital market equilibrium with transaction costs, *Journal of Political Economy*, 94, 842-862.
- Davis, M.H., A.R. Norman, 1990, Portfolio selection with transaction costs, *Mathematics of Operations Research*, 15, 4, 676-713.
- Dixit, A., 1991, *The Art of Smooth Pasting*, Harwood Academic Publishers, Chur, Switzerland.
- Dumas, B., 1991, Super contact and related optimality conditions, *Journal of Economic Dynamics and Control*, 15, 675-685.
- Fleming, W. H., H. M. Soner, 1993, *Controlled Markov Processes and Viscosity Solutions*, New York: Springer Verlag.
- Froot, K.A., D.A. Scharfstein, J.C. Stein, 1993, Risk management: coordinating corporate investment and financing policies, *The Journal of Finance*, 48, 5, 1629-1658.
- Froot, K.A., J.C. Stein, 1998, Risk management, capital budgeting, and capital structure policy for financial institutions: an integrated approach, *Journal of Financial Economics*, 47, 55-82.
- Harrison, J.M., 1985, *Brownian Motion and Stochastic Flow Systems*, Wiley, New York.
- Harrison, J.M., M.Taksar, 1983, Instantaneous control of Brownian motion, *Mathematics of Operations Research*, 8, 439-453.

- Hojgaard, B., M. Taksar, 1999, Controlling risk exposure and dividends payout schemes: insurance company example, *Mathematical Finance*, 9, 153-182.
- Hojgaard, B., M. Taksar, 2001, Optimal risk control for a large corporation in the presence of returns on investments, *Finance and Stochastics*, 5, 527-547.
- Jeanblanc-Pique, M., A. Shiryaev, 1995, Optimization of the flow of dividends, *Russian Mathematical Surveys*, 50, 257-277.
- Karlin, S., Taylor, H. M., 1981, *A Second Course in Stochastic Processes*, Academic Press, New York.
- Mello, A.S., J.E. Parsons, 2000, Hedging and liquidity, *The Review of Financial Studies*, 13, 1, 127-153.
- Milne, A., D. Robertson, 1996, Firm behavior under the threat of liquidation, *Journal of Economic Dynamics and Control*, 20, 1427-1449.
- Milne, A., E. Whalley, 2001, Bank capital regulation and incentives for risk taking, working paper, City University Business School, London.
- Peura, S., J. Keppo, 2002, Optimal bank capital with costly recapitalization, working paper, available at www.ssrn.com.
- Radner, R., 1998, Profit maximization with bankruptcy and variable scale, *Journal of Economic Dynamics and Control*, 22, 849 – 867.
- Radner, R., L. Shepp, 1996, Risk vs. profit potential: a model for corporate strategy, *Journal of Economic Dynamics and Control*, 20, 1373 - 1393.
- Shepp, L., A. Shiryaev, 1996, Hiring and firing optimally in a large corporation, *Journal of Economic Dynamics and Control*, 20, 1523-1539.

Figure 1. The critical value of α

The vertical axis is the critical value of α given by formula (21). Fixed parameter value $\rho = 0.1$. $\hat{\alpha}$ does not depend on δ .

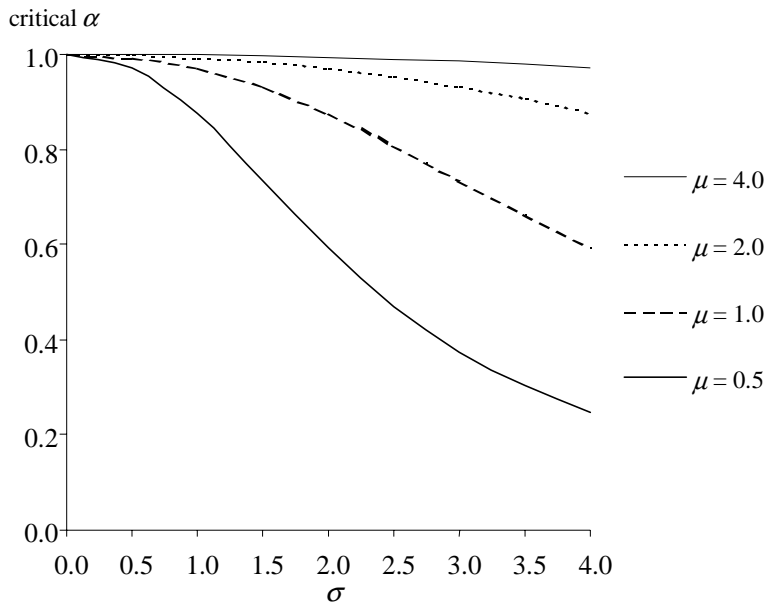


Figure 2. Firm value with (solid line) and without (dotted line) the capital issue option

Horizontal axis: initial capital stock x . Vertical axis: firm value. Fixed parameter values: $\mu = 1$, $\sigma = 2$, $\rho = 0.1$, $\delta = 2$. Variable problem parameter: $\alpha = 0.0$ in (1), 0.2 in (2), 0.4 in (3), and 0.6 in (4).

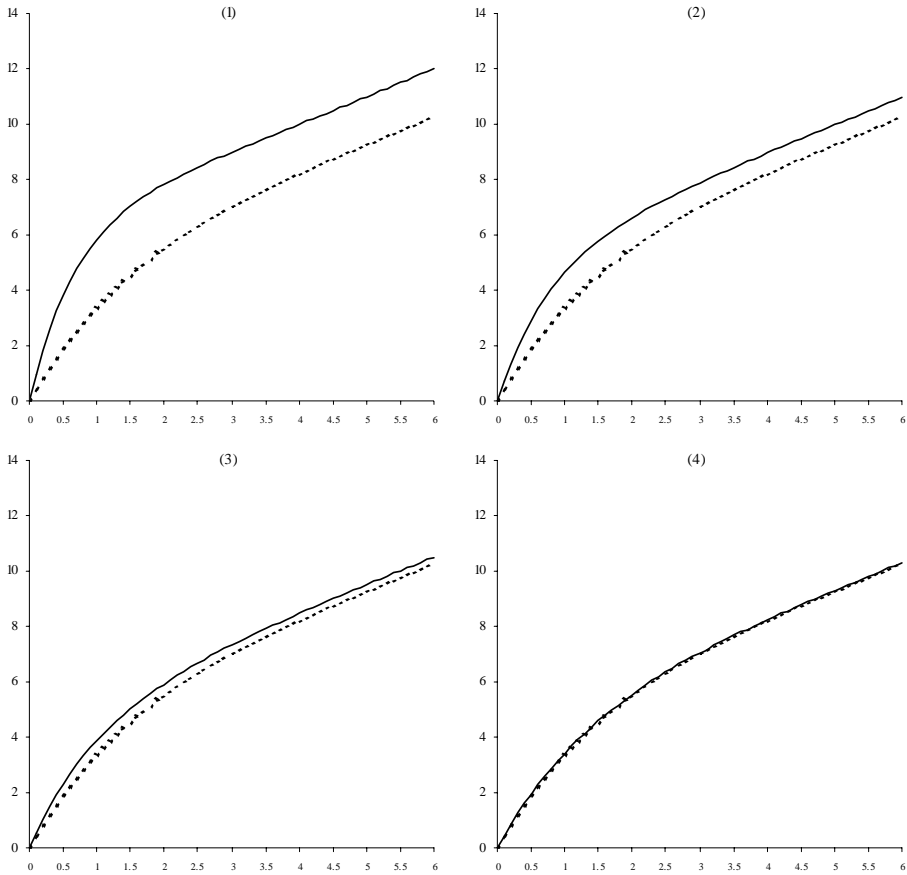


Figure 3. The value of the capital issue option

The upper picture shows the absolute value of the option, the lower picture shows the value of the option as a percentage of the (cum-option) firm value. Fixed parameter values: $\mu = 1$, $\sigma = 2$, $\rho = 0.1$, $\delta = 2$. Given these parameters, b_0 equals 5.74.

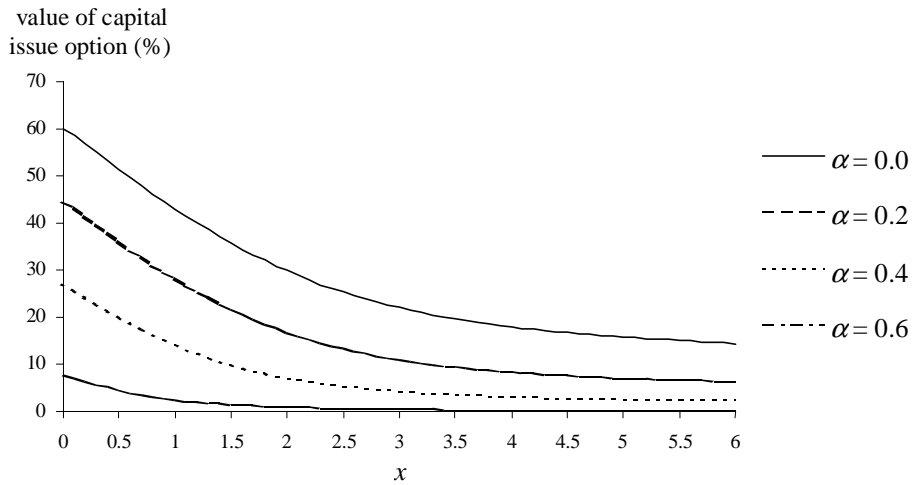
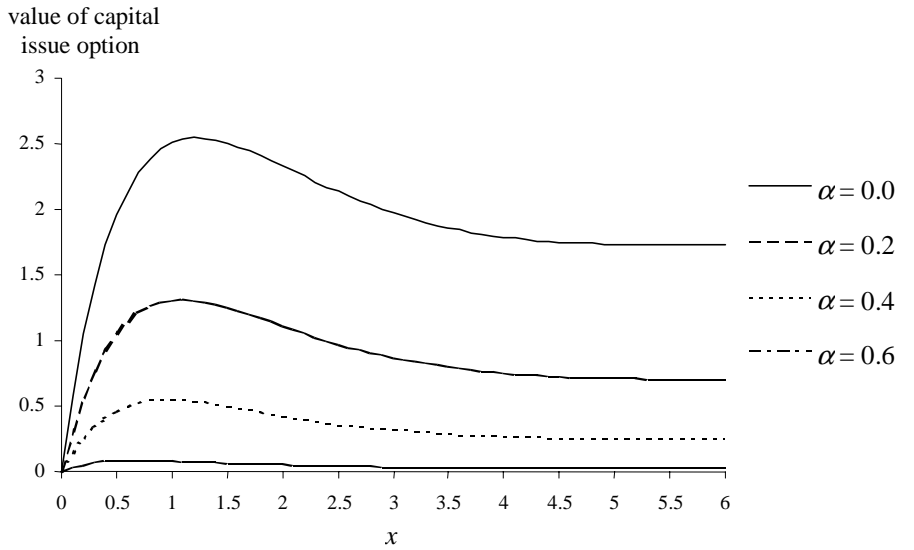


Figure 4. b_1 and b_2 as a function of α

The picture shows the optimal b_1 and b_2 as a function of α , for given values of δ (legend). Fixed parameter values: $\mu = 1$, $\sigma = 2$, $\rho = 0.1$. The critical α given by formula (21) is 0.76. At the critical α , b_1 equals 0, and b_2 equals the dividend barrier b_0 , 5.74.

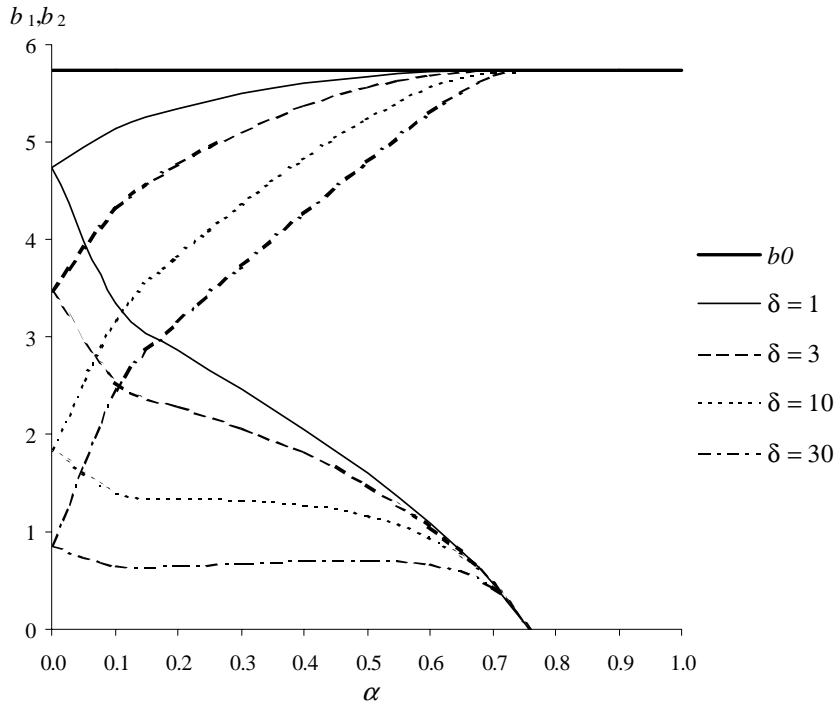


Figure 5. Firm value with the capital issue option (solid line), with the risk reduction option (broken line), and without the options (dotted line)

Horizontal axis: initial capital stock x . Vertical axis: firm value. Fixed parameter values: $\mu = 1$, $\rho = 0.1$, $\alpha = 0.2$. Variable parameters: $\sigma = 1$ in (1) and (2), $\sigma = 2$ in (3) and (4); $\delta = 1$ in (1) and (3), $\delta = 3$ in (2) and (4).

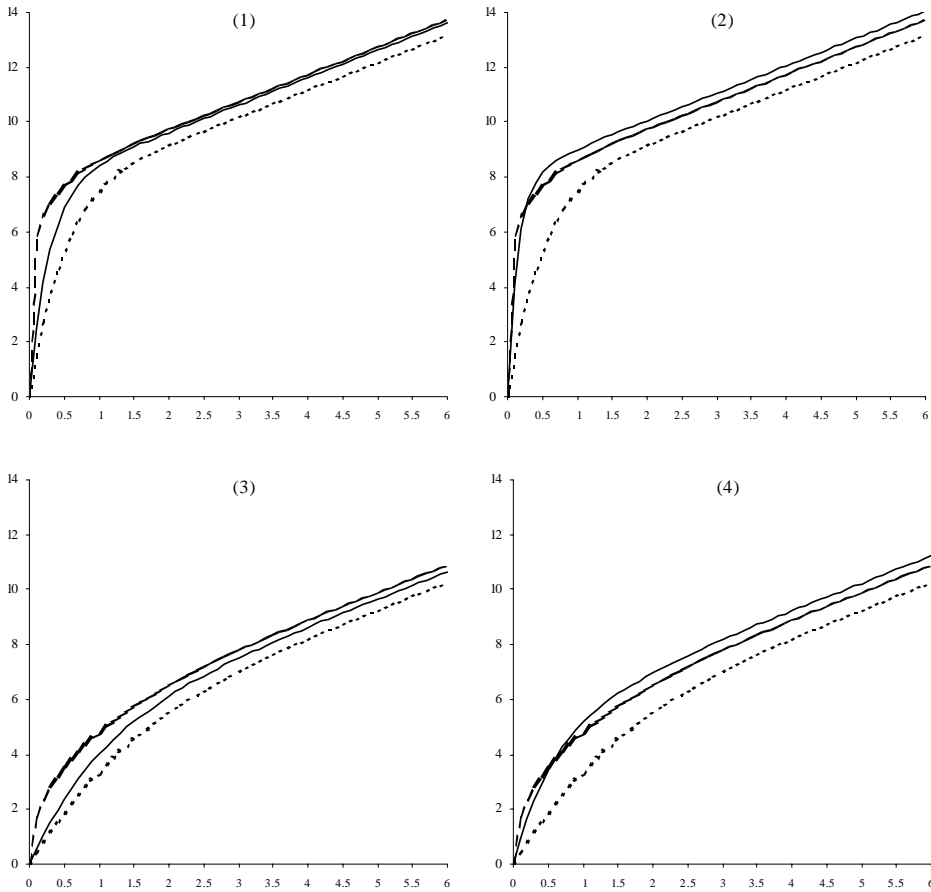
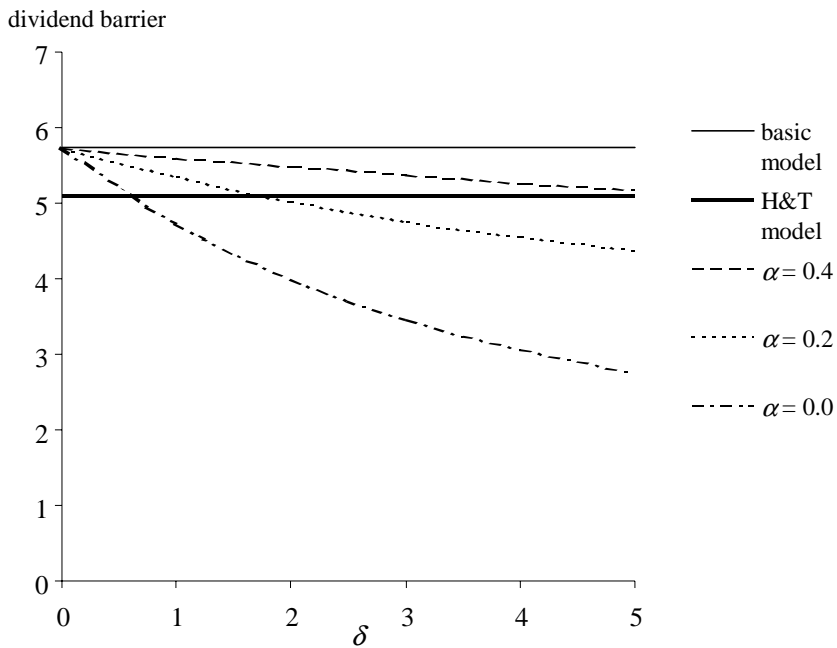


Figure 6. Comparison of dividend payment barriers

The dotted lines are the optimal dividend payment barriers b_2 as a function of the maximal capital injection rate δ (horizontal axis), at three different values of α (legend). The thin solid line is the dividend payment barrier b_0 (28), in the basic model with no capital issuance or risk reduction option, and the thick solid line is the dividend barrier u_2 (34) in the model of Hojgaard and Taksar (1999) with risk reductions. Fixed parameter values: $\mu = 1$, $\sigma = 2$, $\rho = 0.1$.



Essay 2: Optimal bank capital with costly recapitalization

Joint work with Jussi Keppo*

Abstract

We study optimal bank capital holdings in a dynamic setting where the bank has access to external capital, but this access is subject to a fixed cost and a delay. Our model indicates that a recapitalization option may be valuable despite substantial fixed costs, and that a significant fraction of the value of low capitalized banks may be attributable to the option to recapitalize. When calibrated to data on actual bank returns, the model yields capital ratios that are significantly lower than actual bank capital ratios. This shortfall is, at least partly, explained by the skewness of the distribution of actual bank returns and by the banks' accounting options for the provisioning of credit losses. We operate the model with implied bank return volatilities, in the same way as Black-Scholes model is used in practice. Analysis of the limiting cases where the capital market imperfections vanish reveals that the capital issue delay rather than the fixed cost determines the qualitative nature of the solution.

Keywords: bank capital, dividends, capital issues, fixed cost, delay

JEL classification: G32, G35

1 Introduction

A general risk management lesson from models with frictions is that, in the absence of explicit risk management tools such as financial derivatives, firms may choose to hold buffer stocks of liquid assets and capital as hedges against liquidity and earnings risks. The argument for the buffer stock role of liquid assets has been theoretically presented and empirically verified by e.g. Kim et al. (1998) and Opler et al. (1999), who find that firm liquid asset holdings are positively related to cash flow risks. The buffer stock role of equity capital, on the other hand, is supported by many empirical studies on capital structure (e.g. Harris and

* Jussi Keppo is at the Department of Industrial and Operations Engineering, University of Michigan, 1205 Beal Avenue, Ann Arbor, Michigan 48109-2117, USA. E-mail: keppo@umich.edu.

Raviv, 1990, Booth et al., 2001, and Titman and Wessels, 1988), who find that firm leverage is negatively related to earnings volatility. In other words, we observe risk management considerations to influence both corporate investment decisions and corporate financing decisions¹.

In banks the capital structure decision is in its very essence a risk management decision. Capital² in banks is not predominantly a form of financing, but a buffer against asset risks which is needed to ensure the stability of the callable deposit based banking system. Banks rely on high leverage to achieve high equity returns, and operate subject to an explicit minimum capital requirement. As Froot and Stein (1998) have demonstrated, a bank investing in illiquid products may adjust its capital structure in order to accommodate the illiquid risks it chooses to bear. Data on banks confirms that they do hold buffer stocks of capital in excess of their minimum requirements: in the US, commercial banks' median total capital ratio was 12.1% in 2001, and the median total capital ratio of even the largest banks (those with capital in excess of 3 m\$) was over 11% (figures based on Bankscope data). The minimum requirement imposed by the Basel Accord is 8%, implying that the median bank capital buffer among US commercial banks is in excess of 4%. A qualitative explanation for banks' capital holdings could be based on the illiquidity of their asset risks and on some imperfections in their capital raising transactions. There are several models in the banking literature which can deliver predictions on buffer holdings of capital (a review of these models is given below). Yet what appears to be missing from the literature is a calibration of a model of this type to data on actual bank returns and risks, and a comparison of the model predictions against actual bank capital holdings. In this paper we develop a model of a capital constrained bank and perform such calibration.

Our model of a capital regulated bank is an extension of the Milne and Robertson (1996) model of firm dividend policies under the threat of liquidation. In our model, bank equity holders optimize over dividend and equity issuance policies in order to maximize the present value of net distributions from the bank. Capital regulation is characterized by a strict

¹ By now there is a large literature on the interactions between financing, investment, and risk management in the presence of capital market frictions (e.g. Acharya et al., 2000, Froot et al., 1993, Froot and Stein, 1998, Leland, 1998, Mello and Parsons, 2000).

² Consistent with common bank parlance, our use of the term bank capital refers to banks' book equity. This is the relevant measure of capital in an analysis of bank capital adequacy since minimum capital requirements under the Basel Accord apply to book equity.

minimum capital requirement, the violation of which results in liquidation and loss of continuing value. The bank portfolio is completely illiquid and is of constant size (an assumption of constant scale). Because the bank may finance its assets using zero cost (insured) deposits, any capital held in the bank carries an opportunity cost. Optimal capital level is determined from the trade-off between the opportunity cost of holding capital and the expected deadweight costs from liquidation. The bank has access to external (equity) capital, but this access is subject to a fixed cost³ and a delay. These capital market imperfections influence the level of optimal bank capital.

The optimal policies in our model are of the barrier control type. Unless the fixed cost or the delay in capital issues are too high, there is a positive barrier u_1 such that capital issues are ordered at the first time the capital buffer is at or below u_1 . If the bank has not been liquidated by the end of the delay, the owners pay the fixed cost K and raise new capital to the amount that makes aggregate capital buffer equal to a barrier u_2 , which is strictly above u_1 . When capital is between u_1 and u_2 , the bank neither pays dividends nor raises additional capital. Above the barrier u_2 , the bank distributes dividends at maximum admissible rate, which in our model is assumed infinite, effectively preventing the capital buffer from rising above u_2 . Both the barriers u_1 and u_2 are sensitive to the degree of capital market imperfections.

We also solve the limiting cases of our model as the capital market imperfections vanish. When the capital issue delay equals zero, an optimal policy does not exist, but we derive the limiting value function. When the fixed cost from capital issues converges towards zero, the optimal policy does not converge to a degenerate limit, and the optimal policy exists even in the limiting case when the fixed cost equals zero. These results imply that the delay in capital issues, rather than the fixed cost, determines the qualitative nature of the solution. Moreover, as either the fixed cost or the delay are sufficiently high, optimal policies do not involve capital issues and our model reduces to the Milne and Robertson model without the capital issue option.

The defining properties of the Milne and Robertson (1996) model (also analyzed in Asmussen and Taksar, 1997) are the assumption of constant scale in banking activities and

³ The cost of raising capital is likely to be positively related to the amount of capital raised. In section 5.4 of this paper we show how to determine the fixed cost as a function of the other model parameters so as to give a desired expected cost rate in capital issuance.

the objective function which is the present value of dividend distributions. We extend the model by allowing for equity issues, and redefining the objective function as the present value of net distributions. Endogenous determination of bank capital has also been studied using other variants of the same basic model. A closely related paper to ours is Milne and Whalley (2001), where the basic model is extended with Poisson distributed bank capital audits and a fixed cost of recapitalization. In their model, a forced liquidation following bank capital violation is replaced by an option to recapitalize, given an observed violation in a randomly occurring audit. Hojgaard and Taksar (1999) have extended the basic model into an insurance company setting by assuming that risk reduction at a proportional cost is available, which is interpreted as cheap reinsurance. Finally, Peura (2002) has analyzed the effects of an equity issuance option which is subject to a proportional cost, and an upper bound on the rate at which external capital can be raised.

Each of the models listed above predicts that some amount of buffer capital will be optimally held. The models also demonstrate that buffer capital will be the lower, the more effective hedging mechanisms are available or the less expensive is external recapitalization. In fact, optimal capital holdings are the highest in the basic model where no risk reduction or recapitalization options are present, and each of the ‘variants’ will predict lower capital levels. Yet none of the papers has tested this class of models’ ability to explain empirically observed bank capital levels. In this paper we perform a simple calibration of our model (and as its special case the basic model of Milne and Robertson) to empirical data on bank returns, and compare the resulting model capital ratios against observed bank capital ratios. Our findings, stated briefly, indicate that this class of models does not generate sufficiently high capital buffers when calibrated with the empirically observed volatilities of bank returns. The models are able to generate capital ratios of the correct size when volatility is roughly tripled from its empirically observed value. This need for volatility adjustment is partly due to the fact that accounting based estimates of bank return volatility are likely to be downward biased because of banks’ income smoothing options (most notably, the provisioning for expected losses). Some part of this adjustment is also explained by the skew distribution of bank portfolio returns. Actual bank capitalization is likely to reflect the potential for large losses that banks may experience during severe economic downturns, while our model assumes normally distributed asset returns. Our dataset from 1994-2001 indicates that there is negative skewness in banks’ accounting returns, despite the fact that banks’ provisioning policies are likely to

reduce the negative skewness inherent in banks' portfolio returns. Due to this model error, we operate our model with implied bank return volatilities, which are chosen so as to replicate observed bank capital levels. This is analogous to how the Black-Scholes model is used in practise. Our model yields simple quasi-analytic solutions for optimal policies, and can be used to study the effects of the capital market imperfections on optimal capital raising and dividend policies. The model also provides a decomposition of the value of a bank into the value of the capital raising option and the value of a non-optional bank. We illustrate these issues in section 5.

The variants of the Milne and Robertson (1996) model are formulated in continuous time, and rely on stochastic and singular control techniques to obtain optimal policies. The determination of bank capital has also been studied in discrete time by Estrella (2001) and Furfine (2000). Estrella (2001) uses a variant of the classical inventory or cash management models (see e.g. Karlin and Taylor, 1981, pp. 211-212 for the classical cash inventory model). In his model, the objective is to minimize the combined costs from over- and undercapitalization, as well as from adjustment of capital. In Furfine (2000), the objective is the present value of net distributions, net of any adjustment costs. His model is therefore a combination of the inventory theoretic analysis of Estrella (2001) and the present value of dividends optimizing approach on which the variants of the Milne and Robertson (1996) model are based on. Furfine also performs a calibration of his model to panel data on bank capital and profitability, but does not explicitly analyze the absolute capital levels generated by his calibrated model.

Methodologically, our paper is distinguished by the joint presence of an implementation delay and a fixed cost. Both of these frictions have been analyzed in several contexts in the corporate finance literature. Fixed costs from transacting have been analyzed e.g. in the classical inventory control literature, in the literature on optimal portfolio choice with transaction costs (see e.g. Eastham and Hastings, 1988, Korn, 1998, Morton and Pliska, 1995), and in connection to risky debt valuation by Fan and Sundaresan (2000). Fixed costs lead to impulse controls, i.e. controls that change state variables in discrete magnitudes and are applied at discrete stopping times. Impulse control techniques have been elaborated in books by Bensoussan and Lions (1982) and Fleming and Soner (1993). The effects of delays in implementing impulse controls within continuous time models have been analyzed e.g. by Bar-Ilan and Strange (1996) and Alvarez and Keppo (2001) in connection with exercising of

real investment options, and by Subramanian and Jarrow (2001) in connection with optimal liquidation.

The rest of the paper is organized as follows. Section 2 presents our model. The solution is characterized in terms of a set of quasi-variational inequalities in Section 3. Optimal policies and the value function are derived in section 4. Model parameters are calibrated, and results based on the calibrated parameter values are presented in section 5. Section 6 derives the probability of default of an optimally managed bank, and Section 7 concludes with a number of remarks concerning the extendability of the model presented here.

2 The model

We imagine a bank with a fixed portfolio of non-tradable assets. In order to simplify the empirical implementation of the model, we normalize the asset size so that the portfolio's regulatory risk weighted assets (RWA) are equal to unity⁴. The state variable in the model is the bank's (book) capital stock, which is denoted X_t . Since the bank's risk weighted assets are normalized to one, X_t is also the bank's regulatory capital ratio. Bank capital is driven by profits and losses, and in the absence of controls evolves according to

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

where $\{W_t : t \geq 0\}$ is a standard Wiener process, and μ and σ are constant drift and diffusion coefficients, which are both assumed to be positive. We assume that the initial capital stock, X_0 , is nonnegative and generically denoted by x . We also denote the standard filtration generated by the Wiener process $\{W_t : t \geq 0\}$ as $\{F_t : t \geq 0\}$.

Given our normalization $RWA = 1$, we can interpret μ as the bank's expected return on (risk weighted) assets, and σ as the volatility of the bank's return on (risk weighted) assets. That neither μ nor σ depend on the level of bank capital is due to the fact the bank's deposits

⁴ Risk weighted assets, under the current Basel Accord from 1988, are calculated as a weighted sum of the banks' nominal exposures, where the weights depends on product type and counterparty sector. For large banks, risk weighted assets are typically between 65 and 70 percent of total assets. Capital ratio is book equity divided by risk weighted assets.

are assumed to be insured and zero cost (this assumption is motivated in Milne and Whalley, 2001).

The bank's owners control the bank capital ratio through dividend payments and issues of new capital. Formally, a control policy π is a collection $(L^\pi, \{t_i^\pi, s_i^\pi\})$, where L^π is a non-decreasing process representing the cumulative amount of dividends paid under policy π , $\{t_i^\pi\}$ is an increasing sequence of order times of new capital issues, and $\{s_i^\pi\}$ are the amounts of capital raised at each issue of capital. We assume that L^π is a non-decreasing right-continuous process adapted to F_t such that $L_{0-}^\pi = 0$, that each t_i^π is a stopping time of the filtration F_t , and that each s_i^π is measurable with respect to $F_{(t_i^\pi + \Delta)-}$. Additionally, admissible controls satisfy

$$t_{i+1}^\pi - t_i^\pi \geq \Delta \text{ for all } i \geq 1, \quad (2i)$$

$$dL_t^\pi = 0 \text{ for all } t \in (t_i, t_i + \Delta], i \geq 1. \quad (2ii)$$

The positive parameter Δ is the length of the delay associated with a capital issue. When a capital issue is ordered at time t_i , the capital can be raised at time $t_i + \Delta$. The measurability of s_i with respect to $F_{(t_i + \Delta)-}$ means that the owners may decide on the actual amount of capital to be raised at time $t_i + \Delta$ based on all then available information, i.e. they do not need to precommit to any quantity of capital at time t_i when they order the capital issue. Condition (2i) states that a new issue may not be ordered while a previously ordered issue is still waiting to be completed. Condition (2ii) states that dividends may not be paid during the periods between the ordering of a capital issue and the actual capital collection. The condition has important technical merit, but also has an economic justification, in that ruling out dividend payments while a capital issue is under preparation is likely to make the issue more easily marketable, and to help reduce conflicts of incentives between existing and new equity holders. We do not analyze the division of bank value between existing and new equity holders, so the potential incentive conflicts are not explicitly present in our model. We simply think of constraint (2ii) as a restriction set by the capital markets.

The class of admissible policies satisfying the restrictions in (2) is denoted Π . The conditions (2i) and (2ii) together ensure that the capital stock remains uncontrolled during the waiting periods following each order of a new capital issue, and will be solely driven by the diffusion (1). The fact that the diffusion remains uncontrolled during the waiting periods is essential for the tractability of the solution.

The capital stock, as a function of policy π , is denoted X_t^π , and can be expressed in integral form as

$$X_t^\pi = x + \mu t + \sigma W_t - L_t^\pi + \sum_i s_i^\pi \mathbf{I}_{\{t_i^\pi + \Delta \leq t\}}, \quad (3)$$

where $\mathbf{I}_{\{\cdot\}}$ is the indicator function of the event defined in the parenthesis. Dividend payments represent a leakage from the capital stock, while new issues feed to the capital stock.

The minimum capital requirement under the current Basel Accord states that bank capital ratio X_t must at all times exceed 8%. In this paper we assume that the corrective action from violation of the minimum capital requirement will be liquidation⁵. The model bank therefore only operates up to the liquidation time

$$\tau_\pi = \inf\{t : X_t^\pi \leq 8\%\},$$

where X evolves according to (3). However, because the capital dynamics in (1) is independent of the level of capital, we can without loss of generality reinterpret the state variable X_t as the capital ratio in excess of the minimum capital requirement, or the bank's capital buffer. The time of liquidation is then defined as

$$\tau_\pi = \inf\{t : X_t^\pi \leq 0\}. \quad (4)$$

The value of bank under policy π to its owners, given initial level of capital x , is the expected discounted present value of dividends less capital issues until liquidation

⁵ In practice, violation of the minimum capital requirement may not lead to immediate liquidation, but will generate additional costs to the bank due to increased regulatory monitoring. Also the bank's competitive position is likely to be affected. Therefore the bank's owners are likely to lose a significant amount of the bank's economic rent, while our model assumes that all of the economic rent is lost.

$$V_\pi(x) = E \left[\int_0^{\tau_\pi} e^{-\rho t} dL_t^\pi - \sum_i e^{-\rho(t_i + \Delta)} (s_i + K) \mathbf{I}_{\{t_i + \Delta < \tau_\pi\}} \right], \quad (5)$$

where ρ is a positive rate used to discount future periodic cash flows (and determines the size of the opportunity cost from holding of buffer capital), and K is a positive constant representing the fixed costs from a capital issue. The integral in (5) is considered from 0- to τ_π , and hence includes the possible initial dividend payment $L_0^\pi - L_{0-}^\pi$, where we take $L_{0-}^\pi = 0$. The value function of the problem is the value of an optimally managed bank

$$V(x) = \sup_{\pi \in \Pi} V_\pi(x). \quad (6)$$

The problem is to identify $V(x)$ and, if one exists, an admissible policy π^* such that $V(x) = V_{\pi^*}(x)$ for all $x > 0$. The problem has five parameters in total. Of these, μ and σ characterize bank returns, ρ is the discount rate (opportunity cost) associated with bank capital, and Δ and K determine the magnitude of the capital market imperfections.

3 Characterization of optimum

The model is time-homogenous outside the order periods, in which case the current capital stock is sufficient as a state variable. We will characterize the value function to the problem through a set of quasi-variational inequalities. For this purpose, we define the following auxiliary operators.

Let D be the set of real-valued functions on \mathbf{R}_+ . We define the operator $M: D \rightarrow D$ by

$$Mf(x) = E_x \left[e^{-\rho \Delta} \sup_s [f(X_\Delta + s) - s - K] \mathbf{I}_{\{\tau_0 > \Delta\}} \right] \quad (7)$$

for all $x \geq 0$ and $f \in D$, where X_Δ is the value at time Δ of X defined in (1), τ_0 is the first hitting time of 0 of X defined in (1), and the expectation is conditioned on $X_0 = x$. $Mf(x)$ is the expected discounted value of the decision to order new capital when the capital stock is x , given that the ‘continuing value’ of the problem is f . If f is continuous and concave and such

that $u_2 = \inf\{x \geq 0 : f'(x) = 1\} < \infty$, the maximizing s satisfies $f'(X_\Delta + s^*) = 1$. In terms of u_2 , this is $s^* = u_2 - X_\Delta$.

Also we define the infinitesimal generator A associated with the process (1) by

$$Af(x) = \frac{1}{2} \sigma^2 f''(x) + \mu f'(x), \quad (8)$$

for all $x > 0$ and sufficiently regular f . Then, the following characterization of optimum can be established using standard arguments (see e.g. Hojgaard and Taksar, 1999, or Fleming and Soner, 1993).

Proposition 1. *Assume that the value function (6) satisfies Ito's formula. Then it satisfies the following set of inequalities:*

$$\begin{aligned} \text{i)} \quad & V(0) = 0 \\ \text{ii)} \quad & V \geq MV \\ \text{iii)} \quad & (A - \rho)V \leq 0 \\ \text{iv)} \quad & V' \geq 1 \\ \text{v)} \quad & (V - MV)(A - \rho)V(V' - 1) = 0. \end{aligned} \quad (9)$$

(9) is a system of quasi-variational inequalities, which are the first order conditions to our problem and follow from standard dynamic programming arguments applied to the Bellman equation. The first equation follows from 0 being an absorbing state in our model. Inequality (9ii) holds for all x since the value of immediate order of new capital can never exceed the value function by definition of the value function. (9iii) holds since applying no control to the capital stock is always an admissible policy. (9iv) must hold for all x since paying dividends is an admissible policy. (9v) states that in an optimum, one of the inequalities must be tight. That is, for all x either taking no action or taking some of the admissible actions must always represent the optimal policy.

In Proposition 1, we are not assuming that the value function is twice continuously differentiable everywhere. It is well known that the second derivative of a value function in general exhibits a discontinuity at the boundary of the region where impulse control actions

are optimal (see Dumas, 1991). This does not prevent Ito's formula from applying, but the differential generator in inequality (9iii) is to be interpreted in terms of left or right derivatives. Consequently, when we say that a function 'satisfies Ito's formula' we mean that the function is continuously differentiable everywhere, and twice continuously differentiable everywhere except at the boundary of the region where impulse control actions are optimal.

Let f be a solution to (9) which is continuous and concave, and satisfies $u_2 = \inf\{x \geq 0 : f'(x) = 1\} < \infty$. Then an admissible control policy π^* can be constructed recursively from f as follows:

$$\begin{aligned} \text{i)} \quad t_1^{\pi^*} &= \inf\{t \geq 0 : f(X_t^{\pi^*}) = Mf(X_t^{\pi^*})\}, \\ t_{i+1}^{\pi^*} &= \inf\{t \geq t_i^{\pi^*} + \Delta : f(X_t^{\pi^*}) = Mf(X_t^{\pi^*})\}, \quad i \geq 1; \end{aligned} \quad (10)$$

$$\text{ii)} \quad s_i^{\pi^*} = \arg \max_s [f(X_{(t_i^{\pi^*} + \Delta)_-}^{\pi^*} + s') - s' - K] = u_2 - X_{(t_i^{\pi^*} + \Delta)_-}^{\pi^*}, \quad i \geq 1;$$

$$\text{iii)} \quad \{X_t^{\pi^*}, L_t^{\pi^*}\} \text{ solve:}$$

$$X_t^{\pi^*} = x + \mu t + \sigma W_t + \sum_i s_i^{\pi^*} \mathbf{1}_{\{t_i^{\pi^*} + \Delta \leq t\}} - L_t^{\pi^*}$$

$$X_t^{\pi^*} \leq u_2 \quad \text{on } \mathbf{R}_+ \setminus \mathbf{O}$$

$$\int_0^\infty \mathbf{1}_{\{X_t^{\pi^*} < u_2\}} dL_t^{\pi^*} = 0,$$

where $\mathbf{O} = \cup_i (t_i^{\pi^*}, t_i^{\pi^*} + \Delta)$ are the periods between the ordering of capital issues and actual capital collections. The maximum in (ii) exists by the assumptions on f . The interpretation of i) is that every time Mf and f coincide outside order periods, a new capital issue is ordered. ii) states that at every capital collection, just enough capital is raised to shift the capital stock to u_2 . According to iii), dividends are paid so as to never let the capital stock rise above u_2 outside order periods. During order periods, capital stock cannot be controlled.

Appendix A proves the converse to Proposition 1, verifying that the set of inequalities (9) are sufficient conditions for the value function in our problem.

Proposition 2. ('verification theorem') *Let f be a concave solution to (9) which satisfies Ito's formula and is such that $u_2 = \inf\{x \geq 0 : f'(x) = 1\} < \infty$. Let π^* be the admissible policy*

constructed from f according to (10). Then f coincides with the value function in (6) and π^* is the optimal policy which attains f , i.e. $f(x) = V(x) = V_{\pi^*}(x)$.

4 Value function and optimal policies

Constructing a concave solution to (9) requires a guess on its form, i.e. on the order of the 'optimality regions' for each of the policies. We will make a guess that we will verify later in this section to be the correct one. We denote a solution to (9) by f . We assume that there are two boundaries $0 \leq u_1 < u_2 < \infty$ defined by

$$u_1 = \inf\{x \geq 0 : f(x) > Mf(x)\}, \quad (11)$$

$$u_2 = \inf\{x \geq 0 : f'(x) = 1\}, \quad (12)$$

so that f solves (9ii) with equality for $x \leq u_1$, f solves (9iii) with equality for $x \in [u_1, u_2]$, and that f solves (9iv) with equality for $x \geq u_2$. Then f also solves (9i) and (9v) for all $x \geq 0$. In the following, we denote the general solution to (9iii) by f_1 and the general solution to (9iv) by f_2 . The coefficients in the general solutions, as well as the locations of the barriers u_1 and u_2 are to be found from the value matching and smooth pasting conditions at the barriers. In particular, based on standard results (e.g. Dumas, 1991, and Dixit, 1991) we expect f to be continuously differentiable at the impulse control barrier u_1 and twice continuously differentiable at the singular control barrier u_2 . If we can find such a solution to (9), then by Proposition 2 this coincides with the value function of our model.

Stated differently, we assume that the optimal policy is of the following form: i) for $x \in (0, u_1]$, it is optimal to immediately order new capital, ii) for $x \in (u_1, u_2)$, it is optimal neither to order new capital nor to pay dividends, and iii) for $x \in [u_2, \infty)$ it is optimal to pay dividends. Furthermore, we expect to have $u_1 < u_2$ and u_2 finite. u_1 may be 0 when the capital market imperfections are prohibitively high. Figure 1 illustrates the model with this form of solution.

We analyze each of the three regions in turn. For $x \in (u_1, u_2)$, (9iii) holds with equality, so that f satisfies

$$\frac{1}{2}\sigma^2 f''(x) + \mu f'(x) - \rho f(x) = 0. \quad (13)$$

The general solution to (13) is the exponential function

$$f_1(x) = c_1 e^{d_{1+}x} + c_2 e^{d_{1-}x}, \quad (14)$$

$$d_{1\pm} = \frac{1}{\sigma^2} \left[-\mu \pm \sqrt{\mu^2 + 2\rho\sigma^2} \right].$$

As $x \rightarrow u_2$, twice continuous differentiability requires that $f'(u_2^-) = 1$ and that $f''(u_2^-) = 0$.

Substituting these limits into (13) gives

$$f(u_2) = \frac{\mu}{\rho}. \quad (15)$$

For $x > u_2$, we must have $f'(x) = 1$. Solving this first order differential equation with the boundary condition (15) yields

$$f_2(x) = \frac{\mu}{\rho} + (x - u_2), \quad x > u_2. \quad (16)$$

Twice continuous differentiability at u_2 is guaranteed by equating the first and second derivatives of (14) and (16) at u_2 . This determines the coefficients in (14), which then takes the form

$$f_1(x) = a_1 e^{-d_{1+}(u_2-x)} + a_2 e^{-d_{1-}(u_2-x)}, \quad (17)$$

$$a_1 = \frac{d_{1-}}{d_{1+}d_{1-} - d_{1+}^2} > 0, \quad a_2 = \frac{d_{1+}}{d_{1+}d_{1-} - d_{1-}^2} < 0.$$

For $0 \leq x \leq u_1$, (9ii) holds with equality, so that f is determined as the fixed point of the operator M defined in (7). We analyze this operator, assuming that it is applied to a concave function f . As noted before, the supremum within the operator is achieved by $s^* = u_2 - X_\Delta$ for all $x > 0$. Also using (15), we can simplify (7) to

$$Mf(x) = e^{-\rho\Delta} E_x \left[(X_\Delta + \beta) \mathbf{1}_{\{s_0 > \Delta\}} \right] \quad (7')$$

where $\beta = \mu/\rho - K - u_2$ measures the net benefits from new issues of equity. Since β is deterministic, equation (7') further simplifies to

$$\begin{aligned} Mf(x) &= e^{-\rho\Delta} E_{x+\beta} \left[X_{\Delta} \mathbf{I}_{\{\tau_{\beta} > \Delta\}} \right] \\ &= e^{-\rho\Delta} E_{x+\beta} \left[X_{\Delta} (\Delta \wedge \tau_{\beta}) \mathbf{I}_{\{\tau_{\beta} > \Delta\}} \right], \\ &= e^{-\rho\Delta} \int_{\beta}^{\infty} yg(y; \Delta, x + \beta) dy \end{aligned} \quad (7'')$$

where $\tau_{\beta} = \inf\{t : X_t = \beta\}$ and $g(y; \Delta, x + \beta)$ is the density of the absorbed process $X_{\Delta}(\Delta \wedge \tau_{\beta})$ that starts at $x + \beta$. The first equality in (7'') is due to the spatial homogeneity of arithmetic Brownian motion, the second equality follows since the values of X outside the event defined in the indicator function do not affect the expectation, and the third equality is due to the fact that for the absorbed process $X_{\Delta}(\Delta \wedge \tau_{\beta})$, the event in the indicator function is exactly the event that the process has not been absorbed by time Δ . Using the Reflection Principle (see e.g. Borodin and Salminen, 1997), the density can be written as

$$g(y; \Delta, x + \beta) = \varphi(y; \mu\Delta + x + \beta, \sigma\sqrt{\Delta}) - \exp\left(-\frac{2\mu x}{\sigma^2}\right) \varphi(y; \mu\Delta - x + \beta, \sigma\sqrt{\Delta}) \quad (18)$$

where $\varphi(y; \mu, \sigma)$ denotes the density of a normal distribution with mean μ and variance σ^2 , i.e.

$$\varphi(y; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right).$$

By integrating, we get from (7'') and (18) that

$$\begin{aligned} Mf(x) &= e^{-\rho\Delta} \left\{ (x + \mu\Delta + \beta) \Phi\left(\frac{x + \mu\Delta}{\sigma\sqrt{\Delta}}\right) + \sigma\sqrt{\Delta} \varphi\left(\frac{x + \mu\Delta}{\sigma\sqrt{\Delta}}\right) \right. \\ &\quad \left. - e^{-\frac{2\mu x}{\sigma^2}} \left[(-x + \mu\Delta + \beta) \Phi\left(\frac{-x + \mu\Delta}{\sigma\sqrt{\Delta}}\right) + \sigma\sqrt{\Delta} \varphi\left(\frac{-x + \mu\Delta}{\sigma\sqrt{\Delta}}\right) \right] \right\}, \end{aligned} \quad (19)$$

where $\Phi(y)$ and $\varphi(y)$ are the cumulative standard normal distribution and its density. Direct substitution shows that $Mf(0) = 0$. In general, Mf need not be increasing or concave. The derivatives of Mf are quite complicated expressions, and it is difficult to find necessary and

sufficient conditions on the primitive problem parameters that would guarantee e.g. that Mf is increasing or concave. The following sufficient condition, however, proven in Appendix B, is very useful.

Lemma 1. *If $\beta = \mu/\rho - K - u_2 \geq 0$, then $Mf(x) > 0$ and $Mf''(x) < 0$ for all x .*

The importance of Lemma 1 is due to the fact that capital issues cannot be optimal policies in our model unless $\beta \geq 0$. This is evident from (7') because if $\beta < 0$, we have

$$\begin{aligned} Mf(x) &= e^{-\rho\Delta} E_x \left[(X_\Delta + \beta) \mathbf{I}_{\{\tau_0 > \Delta\}} \right] \\ &< e^{-\rho\Delta} E_x \left[X_\Delta \mathbf{I}_{\{\tau_0 > \Delta\}} \right] \leq e^{-\rho\Delta} E_x \left[f(X_\Delta) \mathbf{I}_{\{\tau_0 > \Delta\}} \right] \leq f(x) \end{aligned}$$

where the first inequality is due to $\beta < 0$, the second is due to (9iv), and the third is due to the fact that the expression after the second inequality is the value of doing nothing over the period $[0, \Delta]$. By Lemma 1, we know a priori that if Mf constitutes part of the solution to (9), that part must be concave. This enables us to show that the entire solution to (9) is concave.

The solutions for the functions Mf , f_1 and f_2 in (19), (17) and (16) are in terms of the barrier u_2 . In particular, we observe from the definition of β (see (7')) and from (19) that the function $Mf(x)$ depends on f only through u_2 . Therefore we introduce the notation $M(x, u_2) = Mf(x)$ and $f_1(x, u_2) = f_1(x)$, where $Mf(x)$ is given by (19) and $f_1(x)$ is given by (17). The remaining unknowns are the barriers u_1 and u_2 , which are to be solved from the following value matching and smooth pasting conditions at u_1

$$M(u_1, u_2) = f_1(u_1, u_2) \tag{20i}$$

$$\left. \frac{\partial M(x, u_2)}{\partial x} \right|_{x=u_1} = \left. \frac{\partial f_1(x, u_2)}{\partial x} \right|_{x=u_1} \tag{20ii}$$

This non-linear system of equations is quite complicated algebraically, and closed form solutions for u_1 and u_2 do not exist. Hence ultimately solutions to equation (20) have to be found numerically. The following lemma, proven in Appendix B, provides a sufficient condition under which a positive solution to (20) exists. The condition is expressed in terms of a positive barrier u_0 defined by $f_1(0, u_0) = 0$, which can be solved for

$$u_0 = \frac{2}{d_{1+} - d_{1-}} \ln \left(-\frac{d_{1-}}{d_{1+}} \right). \tag{21}$$

u_0 is the dividend barrier in the model of Milne and Robertson without the capital issue option, and represents an upper bound on the dividend barrier in our model.

Lemma 2. *If $\frac{\partial M(x, u_0)}{\partial x} \Big|_{x=0} > \frac{\partial f_1(x, u_0)}{\partial x} \Big|_{x=0}$, then there exists a solution (u_1, u_2) to (20) satisfying $0 < u_1 < u_2 < u_0$ such that $Mf(x, u_2) \leq f_1(x, u_2)$ for all $0 \leq x \leq u_2$.*

Lemma 2 states that if the derivative of the value of new capital issues at $x = 0$ is sufficiently high, then there will be a positive solution to (20), which implies that the option to issue capital will be optimally exercised for all $x \leq u_1$. Consequently, in cases where a solution to (20) satisfying $0 < u_1 < u_2 < u_0$ exists, we propose the following solution to (9)

$$f(x) = \begin{cases} Mf(x) & 0 \leq x \leq u_1 \\ f_1(x) & u_1 < x < u_2 \\ f_2(x) & u_2 \leq x \end{cases} \quad (22)$$

where Mf is given by (19), f_1 is given by (17) and f_2 is given by (16). The boundaries u_1 and u_2 are the solutions to (20). Appendix B proves the following.

Lemma 3. *Assume that a solution to (20) as described in Lemma 2 exists and that f is defined by (22). Then f is a concave solution to (9) and satisfies Ito's formula.*

Lemma 3 implies that formula (22) satisfies the conditions of Proposition 2, and hence coincides with the value function (6). The solution is also unique by the uniqueness of the value function. This is an indirect proof that a solution to (20) in the region $0 < u_1 < u_2 < u_0$ is unique.

Figure 2 illustrates the solution in this general case. The solution f to (9) coincides with Mf for $x < u_1$, with f_1 for $u_1 < x < u_2$, and with f_2 for $x > u_2$. The functions Mf and f_1 (f_1 and f_2) connect to each other smoothly at the barrier u_1 (u_2). At u_1 , the values and first derivatives of Mf and f_1 coincide, but the second derivatives differ. Mf is more concave at u_1 than f_1 , and the second derivative of the solution at this point experiences a discontinuity. In particular, we have $f''(u_1 +) = f_1''(u_1) > Mf''(u_1) = f''(u_1 -)$. At u_2 , the first and the second derivatives of f_1 and f_2 coincide, and the solution f at u_2 therefore has a continuous second derivative. Mf , f_1 , and f_2 each have continuous first and second derivatives on their entire domains of definition. f therefore possesses a continuous second derivative everywhere except at the point u_1 .

When no solution to (20) in the feasible region exists, the solution to (9) reduces to a special case corresponding to the Milne and Robertson (1996) model without the capital issue option. In this case u_1 equals zero, u_2 equals u_0 given in (21), and the value function reduces to (23) below. A necessary condition for this is

$$\left. \frac{\partial M(x, u_0)}{\partial x} \right|_{x=0} < \left. \frac{\partial f_1(x, u_0)}{\partial x} \right|_{x=0}.$$

Figure 3 illustrates one such case. In this example, the fixed cost K is relatively high so that the benefits from new capital issues, net of their cost, are low. For sufficiently high fixed cost K , it is even possible that Mf has a negative right derivative at zero, in which case Mf will take negative values. We note that Mf need not be globally concave, either. In the following subsection, we will systematically analyze the various limiting cases of our model.

4.1 Limiting cases

In this section, we analyze the limiting cases of the model as either the cost or the delay of capital issues equals zero, or as either of these increases above a critical value. The limiting cases help to understand the comparative statistics of the general model, and show exactly which of the capital market imperfections drive the qualitative results from our model.

Case I: $K \rightarrow \infty$ or $\Delta \rightarrow \infty$

Let V be the value function (6). When $K > \mu/\rho$, so that $\beta < 0$, we have from (7') that

$$MV(x) < e^{-\rho\Delta} E_x [X_\Delta \mathbf{I}_{\{\tau_0 > \Delta\}}] \leq e^{-\rho\Delta} E_x [V(X_\Delta) \mathbf{I}_{\{\tau_0 > \Delta\}}] \leq V(x),$$

for all $x > 0$, where the first inequality follows directly from (7'), the second follows from the fact that the value function satisfies (9iv), and the third holds because the expression after the second inequality is the expected value of doing nothing (taking no controls) during a period of length Δ . Therefore μ/ρ is an upper bound on those K under which ordering of new capital may be an optimal policy for some $x > 0$. Also, we get from (7') that

$$\lim_{\Delta \rightarrow \infty} MV(x) = 0$$

uniformly in x on bounded intervals of the form $[0, u_0]$. This implies that for sufficiently high Δ capital issues are also ruled out as optimal policies.

As either K or Δ are above their critical values, the optimal policies and the value function of our model reduce to those of the Milne and Robertson (1996) model without the capital issue option. The value function is obtained as a special case of (22), by setting $u_1 = 0$ and $u_2 = u_0$ given by (21) (this enforces (9i)). The value function takes the form

$$V(x) = \begin{cases} a_1 e^{-d_{1+}(u_0-x)} + a_2 e^{-d_{1-}(u_0-x)} & x < u_0 \\ \frac{\mu}{\rho} + x - u_0 & x \geq u_0 \end{cases} \quad (23)$$

where d_{1+} and d_{1-} are as in (14), and a_1 and a_2 are as in (17).

Case II: $\Delta = 0$

In the absence of a delay in capital raising, new capital can be issued instantaneously and there is perfect control on the minimum level of capital. Given the opportunity cost of capital in our model, it is clearly optimal to wait until the capital stock falls arbitrarily close to zero before issuing new capital. As zero is an absorbing boundary for the capital stock, however, new issues still have to be implemented before the capital stock actually hits zero. Non-surprisingly, then, an optimal policy in the model without delays does not exist. ε -optimal policies however can be constructed which set the capital issue barrier arbitrarily close to 0. Setting Δ equal to 0, the operator M in (7') simplifies to

$$MV(x) = x + \beta = x + \mu/\rho - u_2 - K,$$

for all $x > 0$, where u_2 is defined as in (12). Taking the limit of this operator, as we let the capital issue point approach zero, we obtain the boundary condition satisfied by the limiting value function in the $\Delta = 0$ case

$$V(0) = \max\left(\lim_{x \rightarrow 0^+} MV(x)\Big|_{\Delta=0}, 0\right) = \max(\mu/\rho - u_2 - K, 0).$$

This boundary condition is to replace (9i). This condition is an endogenous equation in the value function, unlike the original boundary condition $V(0) = 0$, where the right-hand side is exogenous. Yet the solution to this special case is considerably easier to obtain than the solution to the general model, and we derive this in Appendix B. The reason for the relative simplicity of this special case is that instead of two endogenous barriers, the solution in the special case $\Delta = 0$ only has one free barrier. The free barrier whose optimal value we now

denote by \hat{u} represents both the dividend barrier, i.e. the level of the capital stock above which dividends are paid, and the level up to which the capital stock is replenished each time a new capital issue is implemented. The solution is given in the following proposition that is proved in Appendix B.

Proposition 3. *If $K < \mu/\rho - u_0$, where u_0 is given by (21), the value function is*

$$V(x) = \begin{cases} a_1 e^{-d_+ (\hat{u}-x)} + a_2 e^{-d_- (\hat{u}-x)} & x < \hat{u} \\ \frac{\mu}{\rho} + x - \hat{u} & x \geq \hat{u} \end{cases} \quad (24)$$

where $\hat{u} < u_0$ is the unique positive solution for u_2 in the equation

$$a_1 e^{-d_+ u_2} + a_2 e^{-d_- u_2} = \frac{\mu}{\rho} - u_2 - K. \quad (25)$$

Else, the value function and the barrier \hat{u} are identical to (23) and (21).

We note that the parametric form of (24) is the same as that of (23), the difference being the location of the barriers u_0 and \hat{u} . As Proposition 3 shows, $\hat{u} < u_0$ when the condition of the proposition holds, and in this case (24) is a left-shifted version of (23).

We also note that if both Δ and K are equal to 0, then (25) is solved by $\hat{u} = 0$. This limiting case represents perfect market conditions. In perfect markets, no buffer stocks of capital are held, and all profits are immediately paid out as dividends. When losses are realized, the capital to cover the losses is instantaneously raised from capital markets that operate frictionlessly. The controlled capital stock for the firm is a constant at zero.

Case III: $K = 0$

We can say little more about this limiting case than about the general case. The reason is that the limit $K = 0$ does not involve a degeneracy as the limit $\Delta = 0$ does. This is easily seen from the expression (7') for the operator M , where K is present in the β term. Setting K to zero does not influence the qualitative properties of M , and hence the solution for the case $K = 0$ is not qualitatively different from the solution in the presence of positive K . This implies that in a model with delays the presence of a fixed cost is not necessary to ensure the existence of optimum and interior solutions. We will demonstrate this further in the numerical examples on the behavior of the barriers u_1 and u_2 . This is in sharp contrast to the $\Delta = 0$ case, where

optimal policies do not exist, but where the limiting ‘unreachable’ value function reduces to a simpler form as compared to the general case.

5 Calibration and numerical examples

In this section we calibrate the model parameters to data on US banks’ asset returns, and analyze the model’s ability to explain the empirically observed bank capital ratios. We then illustrate optimal policies as functions of the capital market imperfections, and demonstrate how the value of the capital issue option behaves as a function of the bank capital ratio. Finally, we show how to interpret and determine the fixed cost K to be used in our model.

5.1 Calibration of parameters and comparison to actual bank capital ratios

The accounting identity that governs the evolution of bank equity is of the form

$$C_t = C_{t-1} + NI_t - D_t + S_t,$$

where C_t is bank equity (in excess of the minimum capital requirement) at time t , NI_t is net income over period t , D_t is dividends over period t , and S_t is equity issuance over period t . The state variable in our model, X_t , is defined as the bank’s capital ratio (in excess of the minimum requirement 8%), so that the relationship between C_t and X_t is

$$X_t = \frac{C_t}{RWA}.$$

Our model assumes that the bank has constant risk weighted assets, so that a discrete version of the model capital dynamics (3), expressed in terms of accounting variables, becomes

$$\Delta X_t = \frac{NI_t}{RWA} - \frac{D_t}{RWA} + \frac{S_t}{RWA}.$$

This expression and equation (3) suggest that we should interpret the model parameters μ and σ in terms of accounting data as

$$\mu = E\left[\frac{NI}{RWA}\right],$$

$$\sigma = SD\left[\frac{NI}{RWA}\right],$$

where SD denotes standard deviation. For the estimation, we use annual Bankscope data on US commercial banks over the period 1994-2001. The total sample contains 276 banks whose 2001 total assets exceed 1 billion USD. For each bank, we calculate the time-series average of the net income to risk-weighted asset ratio, over the sample period, and the standard deviation of that ratio. Then we take as an estimate of μ the median of the time-series averages across all banks, and as an estimate of σ the median of the volatilities across all banks. We take medians rather than means since we think that medians better describe the typical bank in the data set that includes a small number of banks with exceptionally high mean returns. The resulting estimate for μ is 2.0%, implying that the median US bank's return on risk-weighted assets has averaged 2.0% over the 1994-2001 period. The estimate of σ from the data is 0.5%⁶.

The parameter ρ may be interpreted as the margin between the bank's cost of equity and its deposits. Alternatively, ρ may be interpreted as the earnings-to-price ratio of an optimally capitalized bank within our model, i.e. one whose capital ratio is equal to the dividend barrier u_2 . This follows from (15) which states that the value of the bank (i.e. its market capitalization) in our model is equal to μ/ρ when the capital ratio is at u_2 . We use this interpretation to fix the parameter ρ . In particular, we set ρ equal to 1/15, consistent with a price-to-earnings ratio of 15 for an optimally capitalized bank in our dataset.

The median total capital ratio over the US commercial banks in the Bankscope data is 12.4%. This is the median of the time-series averages of the individual banks' capital ratios over the 1994-2001 period, and we use it as our estimate. The variation in the capital ratios across years is moderate, in that the median capital ratio varies between 11.6% and 12.6% during the period 1994-2001. We are interested in comparing the observed capital ratio to the corresponding prediction from our model. We interpret the dividend barrier u_2 in our model as the target level of buffer capital⁷, and will compare this barrier to the observed average bank capital buffer, which is 12.4% - 8% = 4.4%. Our main comparison is based on Figure 4,

⁶ Had we taken means rather than medians across banks, the corresponding estimates for μ and σ would have been 2.2% and 0.8%, respectively. Hence the ratio of the average volatility to the average drift is somewhat higher than the ratio of the median volatility to the median drift.

⁷ u_2 is an upper bound on capital ratios for optimizing banks in our model, and an average capital ratio in the model would be somewhat below u_2 . Moreover, there is no single obvious definition of average here because the stationary distribution for the capital ratio in our model is a degenerate distribution at 0. This is due to the result (proved in Section 6) that an optimally managed bank in our model bankrupts with probability one in finite time.

which shows the model dividend barrier as a function of the volatility parameter σ , given the calibrated values for the parameters μ and ρ .

The immediate conclusion from Figure 4 is that neither our model, nor the basic model of Milne and Robertson (1996), is able to explain the observed median bank capital ratio when volatility is at its observed value, 0.5%. The model of Milne and Robertson, obtained as a special case of our model through setting Δ equal to infinity, yields capital ratios which are upper bounds on the capital ratios in the presence of the capital issue option, for any given values of the parameters (μ , σ , ρ). The capital buffer generated by the Milne and Robertson model, given the observed bank return volatility of 0.5%, is mere 0.8%, or less than one fifth of the observed value of 4.4%. The presence of a capital issue option with reasonable cost only worsens the fit of the model against empirical data. This conclusion is also independent of the estimate of the fixed cost K , set equal to 1 in Figure 4, since the dividend barrier in the basic model of Milne and Robertson does not depend on K .

The volatility observed in the bank return data, 0.5%, is obviously far too low relative to the average bank return, 2.0%, for our model to generate capital buffers in the range of 4% or above. Bank return volatility in our model would need to be roughly tripled in order to replicate the observed capital ratios. In particular, we observe from Figure 4 that the limiting model of Milne and Robertson yields a capital ratio of 4.4% when evaluated at a volatility of 1.5%, given the calibrated values for the parameters μ and ρ . When the capital issue option is present at reasonably low cost, volatilities in the order of 2% are needed to explain actual bank capital holdings.

We find that there is good reason to expect our model to yield downward biased estimates of bank capital buffers, given the empirically observed volatility. First, our model assumes normally distributed bank returns, while bank portfolio returns are expected to display a skew distribution with a stretched lower tail. This is evident from the results of portfolio models such as CreditMetricsTM (J.P.Morgan, 1997), which simulate bank portfolio returns based on returns on individual counterparties. Also the Bankscope data over the period 1994-2001 indicates that there is negative skewness in bank returns. The median coefficient of skewness across banks in our data is -0.13, even though this dataset hardly contains any bad years in terms of the overall performance of the US banking sector. Moreover, the share of negative returns in our dataset is 1.8%, while the share of negative returns suggested by a normal

distribution with a mean of 2.0% and a volatility of 0.5% is only 0.003%. A volatility of roughly 1% would generate the 1.8% probability of negative returns, given a mean of 2.0%. This is a strong indication that the normal distribution is not a good proxy for bank returns.

Second, and perhaps more importantly, we would expect the volatility of bank returns, estimated from data on banks' net income, to be a downward biased estimate of the true volatility in bank portfolio returns⁸. This is because of the options for the provisioning of credit losses that are available to banks, which allow banks to distribute credit losses that often are realized only during a fraction of quarters over each 'credit cycle', more evenly over time. The smoothing of banks' net income is in fact likely to create both a downward bias in the estimates of volatility of bank returns, and an upward bias in the estimates of skewness of bank returns. The bias in the volatility estimate may explain some of the poor fit of our model, and suggests that the volatility could be scaled up to improve the fit of the model. Unfortunately, determining the correct magnitude of such volatility adjustment is extremely difficult since data on the true timing of banks' losses is not publically available. The bias in the skewness estimate, on the other hand, may explain the low degree of negative skewness observed in our data, and suggests that the normal distribution is even less descriptive of actual bank portfolio returns than does the data on banks' net income indicate⁹.

An improved estimate of true bank returns will be a double edged sword from the perspective of our model. A higher volatility is likely to enhance the empirical fit of our model, but at the same time a larger negative skewness is likely to indicate a more serious departure from the normality assumption. We find that the estimation of banks' true return distributions be an interesting challenge for future research¹⁰. In the absence of such estimates, however, we suggest that our model be used with implied bank return volatilities which correspond to observed bank capital ratios, in an analogous manner as Black-Scholes model and its extensions are used in practice. Our model is analytically quite tractable and easy to implement, and yields many conclusions on the optimal dividend and capital raising strategies, as well as on the behaviour of bank value, as functions of the capital market imperfections. These analyses are illustrated in the next subsections.

⁸ This has been acknowledged in the literature on bank capital, see e.g. Estrella (2001), p. 19.

⁹ Managerial risk aversion is another possible explanation for the high observed capital ratios. This would imply that the bank is managed according to a concave (managers') objective, while our model assumes a linear (owners') objective.

5.2 Comparative statistics with respect to capital market imperfections

Our model predicts that optimal bank capital ratios and optimal bank equity issuance policies are influenced by the degree of capital market imperfections. In figures 5 and 6 we show the response of the barriers u_2 and u_1 to the values of the parameters K and Δ . We have drawn the figures keeping μ and ρ equal to their calibrated values (2% and 1/15, respectively), and we have set σ equal to an implicit value of 1.5%. Based on the results in the previous subsection, this value generates a capital ratio for the median bank that is of the same order of magnitude as the actual median capital ratio.

Figure 5 shows the behavior of the dividend barrier u_2 as a function of K and Δ . Unsurprisingly, we observe that u_2 is non-decreasing with respect to both K and Δ . The optimal dividend barrier is determined by balancing the expected cost of new capital issues as well as the expected loss from liquidation against the time value of delayed dividends. The dividend barrier is non-decreasing relative to the fixed cost since this increases expected capital raising costs. The dividend barrier is non-decreasing relative to the length of the delay since a longer delay, *ceteris paribus*, implies a higher probability of liquidation. We also observe from Figure 5 that the optimal dividend barrier is quite sensitive to the introduction of small fixed costs from capital issuance when the fixed cost is initially zero. When the cost is already sizable, the dividend barrier is relatively insensitive to small increases in that cost.

In the limiting case where Δ equals zero, u_2 is equal to its limiting value \hat{u} given in Proposition 3. This limit is positive when K is positive. When K is below its critical value $\mu/\rho - u_0$, u_2 is increasing in K (this can be easily verified from (25)). When K is at or above this critical value, capital issuance ceases to be an optimal policy, and u_2 then equals u_0 given in (21).

Figure 6 shows the barrier u_1 , the highest order point for new capital issues, as a function of Δ and K . We observe that u_1 is non-increasing with respect to K , for all plotted values of Δ . This is intuitive, since a higher fixed cost reduces the net benefit from capital issuance. However, u_1 does not behave monotonically with respect to Δ . When Δ is relatively low, implying a quick access to new capital, the optimal response to an increase in Δ is to raise u_1 . In this case a longer delay induces ‘earlier’ ordering of new capital, i.e. at a higher level of the

¹⁰ There are other motivations for this task, such as the backtesting of banks’ credit portfolio models.

capital stock. On the other hand, when Δ is relatively high the optimal response to an increase in Δ is to lower the barrier u_1 . This happens because the value of ordering a new capital issue is affected two ways through changes in Δ . First, an increase in Δ increases the probability of liquidation during the delay, *ceteris paribus*, inducing an increase in u_1 . Second, our model forbids dividend payments during the capital issue delay, so that an increase in Δ defers potential dividend payments further into the future, should a capital issue be ordered now, suggesting a decrease in u_1 . It turns out that for low Δ , the former effect dominates, while for sufficiently high Δ , the latter effect dominates. Moreover, Figure 6 suggests that the higher is K , the lower is the point where the positive response of u_1 with respect to Δ turns negative. We also observe that, consistent with Proposition 3, u_1 converges to its limiting value zero as Δ approaches zero.

5.3 Value of the capital issue option

The opportunity to issue new equity, being an option, cannot reduce firm value. The value of the capital issue option in our model is the difference between the value functions (22) and (23), for a given initial capital ratio x . In Figure 7, we show the value of the capital issue option as a function of the initial capital ratio, for different values of the fixed cost K . We make a number of observations. First, the value of the capital issue option is monotonically declining in the fixed cost K (and also in the delay Δ). This is not surprising, since the delay and the fixed cost in our setting are pure business constraints. Second, the value of the capital issue option, as a function of the initial capital ratio, displays a humped shaped behavior that appears to be qualitatively independent of the value of K . The option value is at its highest when the capital ratio is somewhat below the optimal capital order boundary u_1 . When capital is close to u_1 , capital issues both achieve their intended effectiveness and will be used with high probability (indeed with probability 1 when the capital stock is less than u_1). The value of the new issue option decreases as the capital stock falls significantly below the optimal capital order barrier since it becomes increasingly unlikely that the firm is alive when capital becomes available. As the capital stock approaches zero, the value of the capital issue option therefore approaches zero as well. The value of the capital issue option is also reduced when capital increases significantly above u_1 , because there the probability of capital shortages in the near future is low, and the expected present value of the gains from future capital issues is therefore low as well. To conclude, the option to issue new capital is most valuable in

absolute terms to banks that are likely to issue new capital either immediately or in the near future, but are still at a reasonable distance from the point of liquidation.

The examples in Figure 7 indicate that the option to issue capital has some value despite substantial capital market imperfections. In Figure 7 the bank has an annual expected return on risk weighted assets equal to 2%. The option to issue capital still has value when the fixed cost of a capital issue is 3% of the risk weighted assets, i.e. one and a half year's expected profit. Therefore it may pay to recapitalize, even when this means paying out several years' worth of expected earnings in fixed expenses. Also, the value of the capital issue option may constitute a substantial portion of the bank's equity value. In the case where K equals 1% (of the bank's risk-weighted assets) in Figure 7, the value of the capital issue option is at its highest (2.2%) when the firm's capital ratio is 0.6% above the liquidation point. This peak option value is 13% of the corresponding value of a bank which does not possess a capital issue option. Moreover, the value of the new issue option in relative terms (as a percentage of the bank value in the absence of the capital issue option) is monotonically declining in the capital ratio. Hence the proportion of bank value which is attributable to the new issue option is highest with banks that are just above their minimum capital requirement.

5.4 The effective cost of capital issuance

Here we attempt to provide more insights into the interpretation and estimation of the fixed cost K . In practice, we would expect the costs of raising capital to be proportional, or at least positively related, to the size of the equity issue. This in turn suggests that there should be a relation between K and the other model parameters which determine the average size of an equity issue. We find it informative to express the effective cost of equity issuance in our model as the ratio of K to $K+E[s]$. Here $E[s]$ is the expected intake of new capital (notation as in (7)), net of all the costs, so that $K+E[s]$ can be interpreted as the total amount of capital raised, while K is the amount paid in fees and other costs to outsiders. Table 1 presents both the expected size of a new issue, $E[s]$, and the effective cost of a new issue, $K/(K+E[s])$, as a function of the capital market imperfections Δ and K .

We observe from Table 1 that the effective cost of a new issue can be substantial, ranging up to 40% under those circumstances where issues of new capital will be optimally undertaken. Moreover, the effective cost of equity issuance increases with K , but at a lower

rate than K . This is because the expected size of a new issue is also increasing in K . In other words, the bank's owners optimally issue new equity in larger quantities, but less frequently, when K is high, relative to the case where K is low. Such behavior is a natural response to the presence of a fixed cost, and is descriptive of impulse control policies in general. Table 1 also indicates that the increase in the issue size with increasing K is due to both increases in the dividend barrier u_2 and decreases in the capital order barrier u_1 . Finally, we observe that the effective cost of a new issue is quite insensitive to the delay Δ , given a fixed K .

The results in Table 1 are useful in helping to determine the value of K to be used in our model. Let us assume that the management has an opinion on the effective cost rate in capital issuance, as well as an opinion on Δ . Then the implied value of K may be read from (an extension of) Table 1. This provides an alternative way to parameterize our model, one where the degree of capital market imperfections faced by the bank is expressed in terms of the delay Δ and the effective cost rate of equity issuance, given in terms of our notation as $K/(K+E[s])$. In section 5.1 we have already given empirical interpretations to the parameters μ , σ and ρ . On these grounds we believe that the empirical implementation of our model is relatively straightforward. That the model only has five parameters should also be an advantage.

6 The probability of bankruptcy in finite time

Does an optimally managed firm in our setup bankrupt in finite time? The probability of bankruptcy is not explicitly present in the corporate objective, but is determined endogenously as a function of the optimal policies. Previous analysis has shown that value maximization may be consistent with a probability of bankruptcy in finite time that equals one. This is the case in the basic model of Milne and Robertson (1996), in the model with risk choice of Radner and Shepp (1996), and in the model with a recapitalization option of Peura (2002). In the model with proportional risk reductions of Hojgaard and Taksar (1999), on the other hand, bankruptcies may under some parameter combinations be completely eliminated. As for our model, the following lemma is proven in Appendix B.

Lemma 4. *The probability of bankruptcy in finite time equals 1.*

It should be noted that the result depends crucially on the assumption that the capital issue delay Δ is strictly positive. As we have shown, in the limiting case where Δ equals 0 (where an optimal policy does not exist) there are ε -optimal policies under which bankruptcies are zero probability events.

7 Possible extensions

We find that the failure of our model, and of this class of models in general, to explain observed bank capital ratios is most likely attributable to the shared assumption of normally distributed asset returns. The assumption is crucial for analytic tractability, but it would be of great interest to analyze the consequences of replacing that assumption with an asymmetric distribution of bank returns. The asymmetric distribution could correspond in form e.g. to the distributions implied by portfolio models, such as the CreditMetrics (1997) framework. Bank portfolio returns generated in a bottom-up manner from individual asset returns are distinctively asymmetric and long-tailed (relative to the normal distribution) in the presence of positive correlations between individual asset returns. These departures from normality should preferably be taken into account in the solution of the bank's optimization problem, but without sacrificing the dynamic structure of the model and the fully optimizing behavior of the bank. We find that this is an interesting challenge for future research on banks' capitalization and financing decisions.

A second obvious extension were to factor investment considerations, and hence growth possibilities, into our model. The model of constant scale is most convenient analytically, and may provide a reasonable approximation to some banks. Yet it is very likely that a bank that is planning to increase its asset size will choose to hold 'extra' capital temporarily. Similarly, a bank that is expecting its assets to be reduced is likely to factor that expectation into its capitalization decision. The empirical data on banks will contain both banks that are expanding in asset size, as well as banks that are diminishing in asset size. We have taken the median capital ratio of large US banks as a proxy for the capital ratio of a constant scale bank. An extension of our model would take investment options into consideration explicitly. A highly stylized analysis of this type is Sethi and Taksar (2002). A related work based on a discrete time model is Cummins and Nyman (2001). Finally, the assumption that the liquid capital stock does not earn a return may seem problematic. In our model, as in Milne and

Whalley (2001), this was motivated by the perfectly elastic presence of zero cost (insured) deposits. We note that the basic model of Milne and Robertson (1996) has been generalized into this direction by Hojgaard and Taksar (2001), who analyze a model where the capital earns a (possibly stochastic) return, and solve the model explicitly in special cases, such as when the return on the capital equals the risk-free rate (a constant).

References

- Acharya, V., J. Huang, M. Subrahmanyam, R. Sundaram, 2000, Costly financing, optimal payout policies, and the valuation of corporate debt, working paper, New York University.
- Alvarez, L. H. R., J. Keppo, 2002, The impact on delivery lags on irreversible investment under uncertainty, *European Journal of Operational Research*, 136, 173-180.
- Asmussen, S., M Taksar, 1997, Controlled diffusion models for optimal dividend payout, *Insurance: Mathematics and Economics*, 20, 1-15.
- Booth, L., V. Aivazian, A. Demirguc-Kunt, V. Maksimovic, 2001, Capital structures in developing countries, *The Journal of Finance*, 56, 87-130.
- Borodin, A. N., P. Salminen, 1996, *Handbook of Brownian Motion – Facts and Formulae*, Basel: BirkHäuser.
- Bensoussan, A., J. L. Lions, 1982, *Impulse control and quasi-variational inequalities*, Paris: Gauthier-Villars.
- Cummins, J. G., I. Nyman, 2001, Optimal investment with fixed refinancing costs, working paper, Federal Reserve Board.
- Dellacherie, C., P.A., Meyer, 1980, *Probabilites et potentiel. Theorie des martingales*, Paris: Hermann.
- Dumas, B., 1991, Super contact and related optimality conditions, *Journal of Economic Dynamics and Control*, 15, 675-685.
- Eastham, J. F., K. J. Hastings, 1988, Optimal impulse control of portfolios, *Mathematics of Operations Research*, 13, 4, 588-605.
- Estrella, A, 2001, The cyclical behavior of optimal bank capital, working paper, Federal Reserve Bank of New York, New York.
- Fan, H., S. M. Sundaresan, 2000, Debt valuation, renegotiation, and optimal dividend policy, *The Review of Financial Studies*, 13, 4, 1057-1099.
- Fleming, W. H., H. M. Soner, 1993, *Controlled Markov Processes and Viscosity Solutions*, New York: Springer Verlag.

- Froot, K. A., D. S. Scharfstein, J. C. Stein, 1993, Risk management: coordinating corporate investment and financing policies, *The Journal of Finance*, 48, 5, 1629–1658.
- Froot, K. A., J. C. Stein, 1998, Risk management, capital budgeting, and capital structure policy for financial institutions: an integrated approach, *Journal of Financial Economics*, 47, 55-82.
- Furfine, C., 2001, Bank portfolio allocation: the impact of capital requirements, regulatory monitoring, and economic conditions, *Journal of Financial Services Research*, 20, 1, 33-56.
- Harris, M., A. Raviv, 1990, Capital structure and the information role of debt, *The Journal of Finance*, 45, 321-349.
- Højgaard, B., M. Taksar, 1999, Controlling risk exposure and dividends payout schemes: insurance company example, *Mathematical Finance*, 9, 153-182.
- Højgaard, B., M. Taksar, 2001, Optimal risk control for a large corporation in the presence of returns on investments, *Finance and Stochastics*, 5, 527-547.
- J.P.Morgan, 1997, CreditMetrics™ - Technical Document, New York.
- Karlin, S., Taylor, H. M., 1981, *A Second Course in Stochastic Processes*, Academic Press, New York.
- Kim, C.-S., D.C. Mauer, A.E. Sherman, 1998, The determinants of corporate liquidity: theory and evidence, *Journal of Financial and Quantitative Analysis*, 33, 335-359.
- Korn, R., 1998, Portfolio optimization with strictly positive transaction costs and impulse control, *Finance and Stochastics*, 2, 2, 85-114.
- Leland, H., 1998, Agency costs, risk management, and capital structure, *The Journal of Finance*, 53, 4, 1213-1243.
- Mello, A.S., Parsons, J.E., 2000, Hedging and liquidity, *The Review of Financial Studies*, 13, 1, 127 – 153.
- Milne, A., D. Robertson, 1996, Firm behaviour under the threat of liquidation, *Journal of Economic Dynamics and Control*, 20, 1427,1449.
- Milne, A., E. Whalley, 2001, Bank capital regulation and incentives for risk-taking, working paper, City University Business School, London.

- Modigliani, F., M. H. Miller, 1958, The cost of capital, corporation finance, and the theory of investment, *The American Economic Review*, 48, 261-297.
- Morton, A. J., S. R. Pliska, 1995, Optimal portfolio management with fixed transaction costs, *Mathematical Finance*, 5, 4, 337-356.
- Myers, S., N. Majluf, 1984, Corporate financing and investment decisions when firms have information that investors do not have, *Journal of Financial Economics*, 13, 187-221.
- Opler, T., L. Pinkowitz, R. Stulz, R. Williamson, 1999, The determinants and implications of corporate cash holdings, *Journal of Financial Economics*, 52, 3-46.
- Peura, S., 2002, Dividends, costly external capital, and firm value: the case of constant scale, Discussion Paper 536, Department of Economics, University of Helsinki.
- Radner, R., 1998, Profit maximization with bankruptcy and variable scale, *Journal of Economic Dynamics and Control*, 22, 849 – 867.
- Radner, R., L. Shepp, 1996, Risk vs. profit potential: a model for corporate strategy, *Journal of Economic Dynamics and Control*, 20, 1373 - 1393.
- Sethi, S.P., M. Taksar, 2002, Optimal financing of a corporation subject to random returns, *Mathematical Finance*, 12, 2, 155-172.
- Subramanian, A., R. A. Jarrow, 2001, The liquidity discount, *Mathematical Finance*, 11, 4, 447-474.
- Titman. S., R. Wessels, 1988, The determinants of capital structure choice, *The Journal of Finance*, 43, 1-19.

Appendix A: Proof of Proposition 2

We will show the equivalence of a solution to (9) and the value function. First, we utilize the inequalities in (9) and a generalization of Ito's formula to show that a solution to (9) majorizes the value of any serious candidate for an optimal policy, and therefore also the value function. As a second step we show that the value of the policy defined by (10) coincides with the solution to (9). These two steps together imply that a solution to (9) coincides with the value function, and that the policy defined by (10) in terms of the solution to (9) achieves the optimum.

Let f be as assumed in the proposition and choose an admissible policy $(L^\pi, \{t_i^\pi, s_i^\pi\})$ such that $s_i^\pi \leq u_2$ for all i . This restriction is harmless, since optimal policies can never have $X_{t_i+\Delta}^\pi > u_2$ by constraint (2ii). Associated with this policy is the set $\Lambda = \{t : L_t^\pi \neq L_{t-}^\pi\}$. Let $\hat{L}_t^\pi = \sum_{s \in \Lambda, s \leq t} (L_s^\pi - L_{s-}^\pi)$ be the discontinuous part of L^π , and $\tilde{L}_t^\pi = L_t^\pi - \hat{L}_t^\pi$ be the continuous part of L^π . Let \hat{T} be a nonnegative stopping time that satisfies

- i) $T - \Delta \leq \hat{T} \leq T$ for some $T < \infty$
- ii) $\hat{T} \notin \cup_i (t_i^\pi, t_i^\pi + \Delta)$.

It is a consequence of (2i) that such stopping time exists. Then we can write

$$\begin{aligned}
& e^{-\rho(t_j^\pi \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_j^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi\right) - f(x) \\
&= \sum_{i=2}^j \left\{ e^{-\rho(t_i^\pi \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_i^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi\right) - e^{-\rho(t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T}}^\pi\right) \right\} \\
&+ \sum_{i=2}^j \mathbf{I}_{\{t_{i-1}^\pi < \hat{T}\}} \left\{ e^{-\rho(t_{i-1}^\pi + \Delta \wedge \tau_\pi)} f\left(X_{t_{i-1}^\pi + \Delta \wedge \tau_\pi}^\pi\right) - e^{-\rho(t_{i-1}^\pi \wedge \tau_\pi)} f\left(X_{t_{i-1}^\pi \wedge \tau_\pi}^\pi\right) \right\} \\
&+ e^{-\rho(t_j^\pi \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_j^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi\right) - f(x)
\end{aligned} \tag{A1}$$

where we have just decomposed the change in the discounted value of $f(X)$ into those changes that occur during order periods, and outside order periods, respectively. By the generalized Ito formula (Dellacherie and Meyer, 1980), we can express the individual terms in the first row on the right-hand side of (A1) as

$$\begin{aligned}
& e^{-\rho(t_i^\pi \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_i^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi\right) - e^{-\rho(t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T}}^\pi\right) \\
&= \int_{t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T}}^{t_i^\pi \wedge \tau_\pi \wedge \hat{T}} e^{-\rho s} (A - \rho) f(X_s^\pi) ds + \int_{t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T}}^{t_i^\pi \wedge \tau_\pi \wedge \hat{T}} e^{-\rho s} \sigma f'(X_s^\pi) dW_s \\
&- \int_{t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T}}^{t_i^\pi \wedge \tau_\pi \wedge \hat{T}} e^{-\rho s} f'(X_s^\pi) d\tilde{L}_s^\pi + \sum_{s \in \Lambda, t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T} < s \leq t_i^\pi \wedge \tau_\pi \wedge \hat{T}} e^{-\rho s} [f(X_s^\pi) - f(X_{s-}^\pi)]
\end{aligned} \tag{A2}$$

The first term on the right-hand side of (A2) is nonpositive since f satisfies (9iii). Also,

$f(X_s^\pi) - f(X_{s-}^\pi) \leq X_s^\pi - X_{s-}^\pi = -(L_s^\pi - L_{s-}^\pi)$ since $f(x) \geq 1$. Substituting these inequalities into (A2) yields

$$\begin{aligned}
& e^{-\rho(t_i^\pi \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_i^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi\right) \leq e^{-\rho(t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T}}^\pi\right) \\
&- \int_{t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T}}^{t_i^\pi \wedge \tau_\pi \wedge \hat{T}} e^{-\rho s} dL_s^\pi + \int_{t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T}}^{t_i^\pi \wedge \tau_\pi \wedge \hat{T}} e^{-\rho s} \sigma f'(X_s^\pi) dW_s
\end{aligned} \tag{A3}$$

Let $\varepsilon > 0$ and define $\tau_\pi^\varepsilon = \inf\{t : X_t^\pi = \varepsilon\}$. Then (A3) also holds with τ_π replaced by τ_π^ε .

Taking expectation of (A3) yields

$$\begin{aligned}
& E\left[e^{-\rho(t_i^\pi \wedge \tau_\pi^\varepsilon \wedge \hat{T})} f\left(X_{t_i^\pi \wedge \tau_\pi^\varepsilon \wedge \hat{T}}^\pi\right)\right] \\
&\leq E\left[e^{-\rho(t_{i-1}^\pi + \Delta \wedge \tau_\pi^\varepsilon \wedge \hat{T})} f\left(X_{t_{i-1}^\pi + \Delta \wedge \tau_\pi^\varepsilon \wedge \hat{T}}^\pi\right)\right] - E\left[\int_{t_{i-1}^\pi + \Delta \wedge \tau_\pi^\varepsilon \wedge \hat{T}}^{t_i^\pi \wedge \tau_\pi^\varepsilon \wedge \hat{T}} e^{-\rho s} dL_s^\pi\right]
\end{aligned} \tag{A4}$$

where the expectation of the Ito integral is zero since $0 \leq f'(X_s^\pi) \leq f'(\varepsilon)$ by concavity of f .

As $\varepsilon \rightarrow 0$,

$$\int_{t_{i-1}^\pi + \Delta \wedge \tau_\pi^\varepsilon \wedge \hat{T}}^{t_i^\pi \wedge \tau_\pi^\varepsilon \wedge \hat{T}} e^{-\rho s} dL_s^\pi \rightarrow \int_{t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T}}^{t_i^\pi \wedge \tau_\pi \wedge \hat{T}} e^{-\rho s} dL_s^\pi, \tag{A5}$$

where the convergence is monotone from below. Also,

$$e^{-\rho(t_i^\pi \wedge \tau_\pi^\varepsilon \wedge \hat{T})} f\left(X_{t_i^\pi \wedge \tau_\pi^\varepsilon \wedge \hat{T}}^\pi\right) \rightarrow e^{-\rho(t_i^\pi \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_i^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi\right) \tag{A6}$$

$$e^{-\rho(t_{i-1}^\pi + \Delta \wedge \tau_\pi^\varepsilon \wedge \hat{T})} f\left(X_{t_{i-1}^\pi + \Delta \wedge \tau_\pi^\varepsilon \wedge \hat{T}}^\pi\right) \rightarrow e^{-\rho(t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T}}^\pi\right),$$

but the convergence need not be monotone. By concavity of f , we have

$$e^{-\rho(t_i^\pi \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_i^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi\right) \leq a + bX_{t_i^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi \leq k\left(1 + X_{t_i^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi\right)$$

for some positive and finite a , b , and k . Moreover,

$$\begin{aligned} E\left[X_{t_i^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi\right] &= E\left[x + \mu\left(t_i^\pi \wedge \tau_\pi \wedge \hat{T}\right) + \sigma W_{t_i^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi - L_{t_i^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi + \sum_{k=1}^{i-1} s_k^\pi \mathbf{I}_{\{t_k^\pi + \Delta < \tau_\pi \wedge \hat{T}\}}\right] \\ &\leq x + \mu T + (i-1)\mu_2 \end{aligned}$$

where the inequality follows from the definition of \hat{T} and the assumed restriction on the policies π . The same holds also for the second random variable in (A6). Consequently, the random variables on the left-hand side of (A6) are majorized by a positive random variable with finite expectation. Letting $\varepsilon \rightarrow 0$, we can use monotone convergence theorem and dominated convergence theorems, respectively, to all the terms in (A4), yielding

$$\begin{aligned} &E\left[e^{-\rho(t_i^\pi \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_i^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi\right)\right] \\ &\leq E\left[e^{-\rho(t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T}}^\pi\right)\right] - E\left[\int_{t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T}}^{t_i^\pi \wedge \tau_\pi \wedge \hat{T}} e^{-\rho s} dL_s^\pi\right] \end{aligned} \quad (\text{A7})$$

By the exactly same steps, we have for the third row on the right-hand side of (A1)

$$E\left[e^{-\rho(t_i^\pi \wedge \tau_\pi \wedge \hat{T})} f\left(X_{t_i^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi\right)\right] \leq f(x) - E\left[\int_0^{t_i^\pi \wedge \tau_\pi \wedge \hat{T}} e^{-\rho s} dL_s^\pi\right]. \quad (\text{A8})$$

As for the second row on the right-hand side of (A1), we get from the definition of M that

$$\begin{aligned} &e^{-\rho(t_{i-1}^\pi \wedge \tau_\pi)} M f\left(X_{t_{i-1}^\pi \wedge \tau_\pi}^\pi\right) \\ &\geq e^{-\rho(t_{i-1}^\pi \wedge \tau_\pi)} E_{X_{t_{i-1}^\pi \wedge \tau_\pi}^\pi} \left[e^{-\rho \Delta} \left(f\left(X_{t_{i-1}^\pi + \Delta \wedge \tau_\pi}^\pi\right) - s_{i-1}^\pi - K\right) \mathbf{I}_{\{t_{i-1}^\pi + \Delta < \tau_\pi\}}\right] \\ &= E_{X_{t_{i-1}^\pi \wedge \tau_\pi}^\pi} \left[e^{-\rho(t_{i-1}^\pi + \Delta \wedge \tau_\pi)} f\left(X_{t_{i-1}^\pi + \Delta \wedge \tau_\pi}^\pi\right)\right] - E_{X_{t_{i-1}^\pi \wedge \tau_\pi}^\pi} \left[e^{-\rho(t_{i-1}^\pi + \Delta)} (s_{i-1}^\pi + K) \mathbf{I}_{\{t_{i-1}^\pi + \Delta < \tau_\pi\}}\right] \end{aligned} \quad (\text{A9})$$

In (A9), the equality follows from the facts that, first

$$e^{-\rho(t_{i-1}^\pi \wedge \tau_\pi)} e^{-\rho \Delta} \mathbf{I}_{\{t_{i-1}^\pi + \Delta < \tau_\pi\}} = e^{-\rho(t_{i-1}^\pi + \Delta \wedge \tau_\pi)} \mathbf{I}_{\{t_{i-1}^\pi + \Delta < \tau_\pi\}} = e^{-\rho(t_{i-1}^\pi + \Delta)} \mathbf{I}_{\{t_{i-1}^\pi + \Delta < \tau_\pi\}}$$

due to the definition of the indicator function, and second, that the indicator function 'disappears' from the first term after the equality since $f(0) = 0$ by (9i). Substituting (A9) into the second row on the right-hand side of (A1), we get

$$\begin{aligned}
& E_{X_{t_{i-1}^\pi \wedge \tau_\pi}^\pi} \left[e^{-\rho(t_{i-1}^\pi + \Delta \wedge \tau_\pi)} f(X_{t_{i-1}^\pi + \Delta \wedge \tau_\pi}^\pi) - e^{-\rho(t_{i-1}^\pi \wedge \tau_\pi)} f(X_{t_{i-1}^\pi \wedge \tau_\pi}^\pi) \right] \\
& \leq e^{-\rho(t_{i-1}^\pi \wedge \tau_\pi)} Mf(X_{t_{i-1}^\pi \wedge \tau_\pi}^\pi) - e^{-\rho(t_{i-1}^\pi \wedge \tau_\pi)} f(X_{t_{i-1}^\pi \wedge \tau_\pi}^\pi) \\
& + E_{X_{t_{i-1}^\pi \wedge \tau_\pi}^\pi} \left[e^{-\rho(t_{i-1}^\pi + \Delta)} (s_{i-1}^\pi + K) \mathbf{1}_{\{t_{i-1}^\pi + \Delta < \tau_\pi\}} \right] \\
& \leq E_{X_{t_{i-1}^\pi \wedge \tau_\pi}^\pi} \left[e^{-\rho(t_{i-1}^\pi + \Delta)} (s_{i-1}^\pi + K) \mathbf{1}_{\{t_{i-1}^\pi + \Delta < \tau_\pi\}} \right]
\end{aligned} \tag{A10}$$

where the second inequality follows since $f \geq Mf$ by (9ii). Taking expectations of (A1), and substituting in (A7), (A8), and (A10), we get

$$\begin{aligned}
& f(x) - E \left[e^{-\rho(t_j^\pi \wedge \tau_\pi \wedge \hat{T})} f(X_{t_j^\pi \wedge \tau_\pi \wedge \hat{T}}^\pi) \right] \\
& \geq E \left[\int_0^{t_j^\pi \wedge \tau_\pi \wedge \hat{T}} e^{-\rho s} dL_s^\pi \right] + \sum_{i=2}^j \left\{ E \left[\int_{t_{i-1}^\pi + \Delta \wedge \tau_\pi \wedge \hat{T}}^{t_i^\pi \wedge \tau_\pi \wedge \hat{T}} e^{-\rho s} dL_s^\pi \right] \right\} \\
& - \sum_{i=2}^j \mathbf{1}_{\{t_{i-1}^\pi < \hat{T}\}} \left\{ E \left[e^{-\rho(t_{i-1}^\pi + \Delta)} (s_{i-1}^\pi + K) \mathbf{1}_{\{t_{i-1}^\pi + \Delta < \tau_\pi\}} \right] \right\}
\end{aligned} \tag{A11}$$

Since π is admissible, L^π satisfies (2ii). Accounting for this, and letting $j \rightarrow \infty$, (A11) simplifies to

$$\begin{aligned}
& f(x) - E \left[e^{-\rho(\tau_\pi \wedge \hat{T})} f(X_{\tau_\pi \wedge \hat{T}}^\pi) \right] \\
& \geq E \left[\int_0^{\tau_\pi \wedge \hat{T}} e^{-\rho s} dL_s^\pi - \sum_{i \geq 1} \mathbf{1}_{\{t_i^\pi < \hat{T}\}} \left\{ e^{-\rho(t_i^\pi + \Delta)} (s_i^\pi + K) \mathbf{1}_{\{t_i^\pi + \Delta < \tau_\pi\}} \right\} \right].
\end{aligned} \tag{A12}$$

We note in particular that $t_j^\pi \rightarrow \infty$ as $j \rightarrow \infty$ due to condition (2i). Letting $T \rightarrow \infty$, we have

$$E \left[e^{-\rho(\tau_\pi \wedge \hat{T})} f(X_{\tau_\pi \wedge \hat{T}}^\pi) \right] \leq e^{-\rho(T-\Delta)} E \left[k(1 + X_{\tau_\pi \wedge T}^\pi) \right] \rightarrow 0, \tag{A13}$$

where we have again utilized the concavity of f . The convergence above follows from the fact that the expectation of X has a linearly bounded growth rate,

$$E \left[X_{\tau_\pi \wedge t}^\pi \right] \leq x + \mu t + \frac{t}{\Delta} u_2,$$

where we have utilized (2i). Also as $T \rightarrow \infty$,

$$\begin{aligned}
& E \left[\int_0^{\tau_\pi \wedge \hat{T}} e^{-\rho s} dL_s^\pi - \sum_{i \geq 1} \mathbf{I}_{\{t_i^\pi < \hat{T}\}} \left\{ e^{-\rho(t_i^\pi + \Delta)} (s_i^\pi + K) \mathbf{I}_{\{t_i^\pi + \Delta < \tau_\pi\}} \right\} \right] \\
& \rightarrow E \left[\int_0^{\tau_\pi} e^{-\rho s} dL_s^\pi - \sum_{i \geq 1} \left\{ e^{-\rho(t_i^\pi + \Delta)} (s_i^\pi + K) \mathbf{I}_{\{t_i^\pi + \Delta < \tau_\pi\}} \right\} \right]
\end{aligned} \tag{A14}$$

where the convergence is due to monotone convergence on the part of the first term in (A14). On the part of the second term, the convergence follows from the boundedness from above of s_i^π and K , and from (2i). Combining (A12), (A13), and (A14), we have shown that

$$f(x) \geq V_\pi(x). \tag{A15}$$

Since π is arbitrary among the class of policies which may be candidates for optimal policies, we have that $f(x) \geq V(x)$. On the other hand, if we take the admissible policy π^* defined in (10), we can go through the same steps as above and we will get (A15) with equality. In particular, we have equality in (A3) because of (10iii) (dividends are never paid below the region where $f(x)$ equals 1) and the 'complementary slackness' condition (9v) that f must satisfy (f therefore satisfies (9iii) with equality in the region of 'inaction'). We have equality in (A9) due to (10ii), and equality in (A10) due to (10i). These imply that $f(x) = V_{\pi^*}(x) \leq V(x)$. Combining this and (A15), we have $f(x) = V_{\pi^*}(x) = V(x)$. End of proof.

Appendix B: Other proofs

Auxiliary lemma 1 (this will be needed in the proofs of lemmas 1 and 2).

$$E_x[X_\Delta \mathbf{I}_{\{\tau_0 > \Delta\}}] = x + \mu\Delta - \mu \int_0^\Delta p(x, t) dt,$$

where $p(x, t) = P[\tau_0 \leq t | X_0 = x]$.

Proof:

$$\begin{aligned} E_x[X_\Delta \mathbf{I}_{\{\tau_0 > \Delta\}}] &= E_x[X_\Delta \mathbf{I}_{\{\Omega\}}] - E_x[X_\Delta \mathbf{I}_{\{\tau_0 \leq \Delta\}}] \\ &= E_x[X_\Delta] - \int_0^\Delta E[X_\Delta | X_t = 0] \frac{\partial}{\partial t} p(x, t) dt \\ &= x + \mu\Delta - \int_0^\Delta \mu(\Delta - t) \frac{\partial}{\partial t} p(x, t) dt \\ &= x + \mu\Delta(1 - p(x, \Delta)) + \mu \int_0^\Delta t \frac{\partial}{\partial t} p(x, t) dt \\ &= x + \mu\Delta - \mu \int_0^\Delta p(x, t) dt \end{aligned}$$

The first equality follows from the linearity of the expectation, the second from the law of total probability, the third from the Strong Markov property of arithmetic Brownian motion, the fourth just rearranges, and the fifth follows from integration by parts. End of proof.

Lemma 1. *If $\beta = \mu/\rho - K - u_2 \geq 0$, then $Mf(x) > 0$ and $Mf''(x) < 0$ for all x .*

Proof: Let us rewrite the operator M starting from (7) as follows

$$\begin{aligned} Mf(x) &= e^{-\rho\Delta} E_x[(\beta + X_\Delta) \mathbf{I}_{\{\tau_0 > \Delta\}}] \\ &= e^{-\rho\Delta} \left\{ \beta(1 - p(x, \Delta)) + E_x[X_\Delta \mathbf{I}_{\{\tau_0 > \Delta\}}] \right\} \\ &= e^{-\rho\Delta} \left\{ \beta(1 - p(x, \Delta)) + x + \mu\Delta - \mu \int_0^\Delta p(x, t) dt \right\} \end{aligned} \tag{B1}$$

where again $p(x, t) = P[\tau_0 \leq t | X_0 = x]$ and the third equality utilizes the Auxiliary lemma 1.

We know that p satisfies the Kolmogorov backward equation, and that its partial derivatives satisfy

$$\frac{\partial}{\partial x} p(x,t) < 0, \quad \frac{\partial^2}{\partial x^2} p(x,t) > 0, \quad \frac{\partial}{\partial t} p(x,t) > 0,$$

for all $(x,t) \in \mathbf{R}_{++} \times \mathbf{R}_{++}$. We differentiate the final expression in (B1) once and twice with respect to x and obtain

$$\begin{aligned} Mf'(x) &= e^{-\rho\Delta} \left\{ -\beta \frac{\partial p(x,\Delta)}{\partial x} + 1 - \mu \int_0^\Delta \frac{\partial p(x,t)}{\partial x} dt \right\} > 0 \\ Mf''(x) &= e^{-\rho\Delta} \left\{ -\beta \frac{\partial^2 p(x,\Delta)}{\partial x^2} - \mu \int_0^\Delta \frac{\partial^2 p(x,t)}{\partial x^2} dt \right\} < 0, \end{aligned} \quad (\text{B2})$$

where the inequalities hold for all non-negative β because of the signs of the partials of $p(x,t)$. End of proof.

Auxiliary lemma 2. Let $K \geq 0$. For all $0 < u_2 \leq u_0$, $Mf(x) < f_2(x) = \frac{\mu}{\rho} + x - u_2$ for all $x \geq u_2$.

Proof. We use here the notation $Mf(x) = Mf(x, u_2)$ and $f_2(x) = f_2(x, u_2)$. Beginning with (7'), we get

$$\begin{aligned} Mf(x, u_2) &= e^{-\rho\Delta} E_x \left[(X_\Delta + \beta) \mathbf{I}_{\{\tau_0 > \Delta\}} \right] \\ &= e^{-\rho\Delta} E_x \left[\left(X_\Delta + \frac{\mu}{\rho} - K - u_2 \right) \mathbf{I}_{\{\tau_0 > \Delta\}} \right] \\ &= e^{-\rho\Delta} \left\{ E_x \left[X_\Delta \mathbf{I}_{\{\tau_0 > \Delta\}} \right] + E_x \left[\left(\frac{\mu}{\rho} - K - u_2 \right) \mathbf{I}_{\{\tau_0 > \Delta\}} \right] \right\} \\ &= e^{-\rho\Delta} \left\{ x + \mu\Delta - \mu \int_0^\Delta p(x,t) dt + \left(\frac{\mu}{\rho} - K - u_2 \right) (1 - p(x, \Delta)) \right\} \\ &\leq e^{-\rho\Delta} \left\{ \frac{\mu}{\rho} + \mu\Delta + x - u_2 - \mu \int_0^\Delta p(x,t) dt - p(x, \Delta) \left(\frac{\mu}{\rho} - u_2 \right) \right\} \\ &< e^{-\rho\Delta} \left\{ \frac{\mu}{\rho} + \mu\Delta + x - u_2 \right\} \\ &< \frac{\mu}{\rho} + x - u_2 = f_2(x, u_2) \end{aligned}$$

where $p(x,t)$ is as defined in Auxiliary lemma 1. The fourth equality utilizes Auxiliary lemma 1, the first inequality is due to setting K to 0, the second inequality is because $u_2 \leq u_0 < \mu/\rho$ (this is due to (15) and (9iv)), the third inequality is because

$$e^{-\rho\Delta} \left\{ \frac{\mu}{\rho} + \mu\Delta \right\} = e^{-\rho\Delta} \left\{ \int_0^{\infty} e^{-\rho t} \mu dt + \mu\Delta \right\} < \int_0^{\infty} e^{-\rho t} \mu dt = \frac{\mu}{\rho},$$

and the final equality is due to (16). End of proof.

Lemma 2. *If $\frac{\partial M(x, u_0)}{\partial x} \Big|_{x=0} > \frac{\partial f_1(x, u_0)}{\partial x} \Big|_{x=0}$, then there exists a solution (u_1, u_2) to (20) satisfying $0 < u_1 < u_2 < u_0$ such that $Mf(x, u_2) \leq f_1(x, u_2)$ for all $0 \leq x \leq u_2$.*

Proof. We use here the notation $Mf(x) = Mf(x, u_2)$ and $f_1(x) = f_1(x, u_2)$. Throughout, we suppose that the condition $\frac{\partial M(x, u_0)}{\partial x} \Big|_{x=0} > \frac{\partial f_1(x, u_0)}{\partial x} \Big|_{x=0}$ holds.

i) From (17) we get that $\frac{\partial f_1(x, u_2)}{\partial u_2} < 0$, while in (21) we have defined u_0 such that $f_1(0, u_0) = 0$. Therefore $Mf(0, u_0) = 0 = f_1(0, u_0)$, while $Mf(0, u_2) = 0 < f_1(0, u_2)$ for $0 < u_2 < u_0$. On the other hand, by Auxiliary lemma 2, $Mf(u_2, u_2) < f_1(u_2, u_2) = \frac{\mu}{\rho}$ for $0 < u_2 \leq u_0$. The previous facts imply that there is a positive x within $(0, u_0)$ such that $Mf(x, u_0) > f_1(x, u_0)$, and that in general the crossing between $Mf(x, u_0)$ and $f_1(x, u_0)$ is not smooth.

ii) From (19) and (17) we get (given positive Δ)

$$\lim_{\substack{u_2 \rightarrow 0 \\ 0 < x < u_2}} Mf(x, u_2) = 0 < \frac{\mu}{\rho} = \lim_{\substack{u_2 \rightarrow 0 \\ 0 < x < u_2}} f_1(x, u_2).$$

Based on Lemma 1, if $\mu/\rho - K - u_2 \geq 0$ then $Mf(x, u_2)$ is increasing and concave in x , while from (17) we get that for $x < u_2$ also $f_1(x, u_2)$ is increasing and concave in x . This, together with the previous inequality, implies that one can always find a (sufficiently small) u_2 satisfying $0 < u_2 < u_0$ such that $Mf(x, u_2) < f_1(x, u_2)$ for $0 \leq x \leq u_2$.

iii) Based on i) and ii), and the continuity of $Mf(x, u_2)$ and $f_1(x, u_2)$ with respect to x and u_2 , there will also exist a u_2 satisfying $0 < u_2 < u_0$ such that for some $0 < u_1 < u_2$, $Mf(u_1, u_2) = f_1(u_1, u_2)$, while $Mf(x, u_2) \leq f_1(x, u_2)$ for all $0 \leq x \leq u_2$. (By Auxiliary lemma 2, we then also have $Mf(x, u_2) < f_2(x, u_2)$ for all $x > u_2$). But at this choice of (u_1, u_2) , continuous differentiability of $Mf(x, u_2)$ and $f_1(x, u_2)$ with respect to x also implies that

$\frac{\partial Mf(x, u_2)}{\partial x} \Big|_{x=u_1} = \frac{\partial f_1(x, u_2)}{\partial x} \Big|_{x=u_1}$. This is because two continuously differentiable functions whose values coincide in the interior of their domain, but which do not cross, must possess equal derivatives at the point. End of proof.

Lemma 3. *Assume that a solution to (20) as described in Lemma 2 exists and that f is defined by (22). Then f is a concave solution to (9) and satisfies Ito's formula.*

Proof: f is concave: By construction $f'(x) = 0$ for $x \geq u_2$. Differentiating f_1 given by (17) three times shows that $f''' > 0$ on (u_1, u_2) . Therefore f has an increasing second derivative on (u_1, u_2) , which combined with the fact that $f_1''(u_2) = 0$ implies that $f'(x) < 0$ on (u_1, u_2) . Finally, we know from Lemma 1 that Mf is globally concave, and therefore f is concave on $(0, u_1)$. Equality of first derivatives of Mf and f_1 at u_1 then implies that f is globally concave.

f satisfies Ito's formula because each of the component solutions is twice continuously differentiable, while f satisfies the smooth pasting conditions at the barriers u_1 and u_2 .

f solves (9):

(9i): $f(0) = 0$ because $Mf(0) = 0$.

(9ii): $f(x) = Mf(x)$ for $0 \leq x \leq u_1$ by construction. By Lemma 2, $f(x) = f_1(x) \geq Mf(x)$ for $0 \leq x \leq u_2$, and by Auxiliary lemma 2 $f(x) = f_2(x) > Mf(x)$ for $x \geq u_2$.

(9iii): For $0 < x < u_1$, we have $f(x) = Mf(x)$, and we get from Itô's formula

$$e^{-\rho\tau} Mf(X_\tau) = Mf(x) + \int_0^\tau (A - \rho) Mf(X_s) ds + \int_0^\tau \frac{\partial Mf(X_s)}{\partial X_s} \sigma dW_s,$$

where τ is a stopping time defined by $\tau = \tau_\varepsilon \wedge \varepsilon$, for some $\varepsilon > 0$ such that $x - \varepsilon > 0$, $x + \varepsilon < u_1$, and $\tau_\varepsilon = \inf\{t \geq 0 : X_t \notin (x - \varepsilon, x + \varepsilon) | X_0 = x\}$. Taking expectations and noting that the last term is a martingale because $Mf(x)$ is concave, we obtain

$$E_x \left[e^{-\rho\tau} Mf(X_\tau) \right] = Mf(x) + E_x \left[\int_0^\tau (A - \rho) Mf(X_s) ds \right],$$

where $E_x \left[e^{-\rho\tau} Mf(X_\tau) \right]$ is the value of waiting until τ prior to ordering a new capital issue. Because immediate ordering of capital is the optimal action at $x < u_1$, we have

$$Mf(x) \geq E_x \left[e^{-\rho\tau} Mf(X_\tau) \right].$$

Combining the last two equations and the fact that $f(x) = Mf(x)$ for $x < u_1$, we get

$$E_x \left[\int_0^\tau (A - \rho) f(X_s) ds \right] \leq 0.$$

A limit operation then gives us

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{E_x[\tau]} E_x \left[\int_0^\tau (A - \rho) f(X_s) ds \right] = (A - \rho) f(x) \leq 0,$$

for $x < u_1$.

For $u_1 \leq x \leq u_2$, we have $(A - \rho)f(x) = 0$ by construction.

For $x > u_2$, we have $f(x) = \mu/\rho + x - u_2$, $f'(x) = 1$, $f''(x) = 0$, so that

$$(A - \rho)f(x) = \frac{1}{2} \sigma^2 f''(x) + \mu f'(x) - \rho f(x) = \mu - \rho \left(\frac{\mu}{\rho} + x - u_2 \right) = u_2 - x < 0$$

for all $x > u_2$.

(9iv): $f'(x) = 1$ for $x \geq u_2$ by construction. That $f'(x) > 1$ for $x < u_2$ follows from the concavity of f (proved above).

(9v): Follows directly from construction.

End of proof.

Proof of Proposition 3. When $\Delta = 0$, we get from (7') that $Mf(x) = x + \mu/\rho - u_2 - K$, for all $x > 0$, where u_2 is the dividend barrier defined in (12). The boundary condition that replaces (9i) is given by the maximum of $\lim_{x \rightarrow 0^+} Mf(x) = \mu/\rho - u_2 - K$ and zero. With this modification, the set of quasi-variational inequalities in (9) becomes

$$f(0) = \max(\mu/\rho - u_2 - K, 0)$$

$$(A - \rho)f \leq 0$$

$$f' \geq 1$$

$$(A - \rho)f(f' - 1) = 0.$$

Solving ii) and iii) as previously, and enforcing smooth pasting at u_2 , yields expressions (16) and (17) in the text. The barrier u_2 is solved by substituting (17) into the boundary condition i). Assuming that the maximum in i) is achieved by the first term, u_2 is determined from the equation

$$f(0) = g(u_2) = a_1 e^{-d_1 u_2} + a_2 e^{-d_1 u_2} = \frac{\mu}{\rho} - u_2 - K. \quad (\text{B3})$$

We record the following easily verifiable properties of the function $g: \mathbf{R}_+ \rightarrow \mathbf{R}$ defined in (B3): i) $g(0) = \mu/\rho$, ii) $g'(u) < 0, u \geq 0$, iii) $g(u) \rightarrow -\infty$ as $u \rightarrow \infty$, iv) $g''(0) = 0$, v) $g'''(u) < 0, u \geq 0$, where the convergence in iii) is exponential. Since K is positive, we have $g(0) = \frac{\mu}{\rho} > \frac{\mu}{\rho} - K$. g is concave by iv) and v), and therefore crosses the linear function $\mu/\rho - u_2 - K$ at a unique $\hat{u} > 0$ which satisfies (B3).

For the maximum in i) to be achieved by the first term, the solution to (B3) must satisfy

$$f(0) = g(\hat{u}) = \frac{\mu}{\rho} - \hat{u} - K > 0, \quad (\text{B4})$$

i.e. $g(u_2)$ must be positive at the point of intersection with the function $\mu/\rho - u_2 - K$. Setting $g(u) = 0$, we get that $u = u_0$, where u_0 is given by equation (21) in the text. Then a necessary and sufficient condition for (B4) to hold is that $\mu/\rho - u_0 - K > 0$, or equivalently that $K < \mu/\rho - u_0$, which is the condition in the proposition. Since g is declining in u , in this case also $\hat{u} < u_0$. End of proof.

Lemma 4. *The probability of bankruptcy in finite time is 1.*

Proof: We only need to analyze the case $u_1 > 0$ since the case $u_1 = 0$ (the basic model in the absence of capital issues) the result is known to hold. When $u_1 > 0$, default can occur only during the capital issue's time delay. Capital issue is ordered when $X_t \leq u_1$ and, because $u_1 > 0$, we get from (7'') that the probability of no default during Δ is given by

$$P[X_\Delta(\Delta \wedge \tau_0) > 0] = \int_0^\infty g(y; \Delta, u_1) dy < 1$$

where $g(y; \Delta, u_1) = \varphi(y; \mu\Delta + u_1, \sigma\sqrt{\Delta}) - \exp(-\frac{2\mu u_1}{\sigma^2})\varphi(y; \mu\Delta - u_1, \sigma\sqrt{\Delta})$, and $\varphi(y; \mu, \sigma)$ is the density of a normal distribution with mean μ and variance σ^2 , i.e.

$$\varphi(y; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right).$$

The above equation implies that the probability of default during the delay is $P[X_\Delta(\Delta \wedge \tau_0) = 0] < 1$. The probability of capital issue on $[0, T]$ is bounded below as follows

$$\begin{aligned} P[X_T(T \wedge \tau_{u_1}) \leq u_1] &\geq 1 - \int_{u_1}^{\infty} g(y; T, u_2) dy \\ &= 1 - \int_{u_1}^{\infty} \varphi(y; \mu T + u_2 + u_1, \sigma\sqrt{T}) dy + \exp\left(-\frac{2\mu u_2}{\sigma^2}\right) \int_{u_1}^{\infty} \varphi(y; \mu T - u_2 + u_1, \sigma\sqrt{T}) dy \quad (B5) \\ &= 1 - \frac{1}{\sigma\sqrt{T}} \left[\Phi(\mu T + u_2) + \exp\left(-\frac{2\mu u_2}{\sigma^2}\right) \Phi(\mu T - u_2) \right] \end{aligned}$$

We have the inequality because on the right-hand side the dividends are not considered and the process starts at u_2 . Now taking limit we get $\lim_{T \rightarrow \infty} P[X_T(T \wedge \tau_{u_1}) \leq u_1] = 1$ because

$$\Phi(\mu T + u_2) + \exp\left(-\frac{2\mu u_2}{\sigma^2}\right) \Phi(\mu T - u_2) \leq 2.$$

From (B5) we also get that the capital stock without capital issues satisfies

$$\lim_{T \rightarrow \infty} P[X_{T \wedge u_1 - (n-1)(u_2 - u_1)}^\pi \leq u_1 - (n-1)(u_2 - u_1)] = 1 \quad \text{for all } n \in \{1, 2, \dots\},$$

where $X_{T \wedge u_1 - (n-1)(u_2 - u_1)}^\pi = x + (T \wedge \tau_{u_1 - (n-1)(u_2 - u_1)})\mu + \sigma W_{T \wedge u_1 - (n-1)(u_2 - u_1)} - L_{T \wedge u_1 - (n-1)(u_2 - u_1)}^\pi$ and n indicates the number of capital issues we should have done. This implies that the capital stock process with capital issues hits u_1 infinite often. Therefore, the probability of no default over infinite time horizon is given by $\lim_{n \rightarrow \infty} P[X_\Delta(\Delta \wedge \tau_0) > 0]^n = 0$ because $P[X_\Delta(\Delta \wedge \tau_0) > 0] < 1$.

The company has finite default time with probability 1. End of proof.

Figure 1. Illustration of the model structure

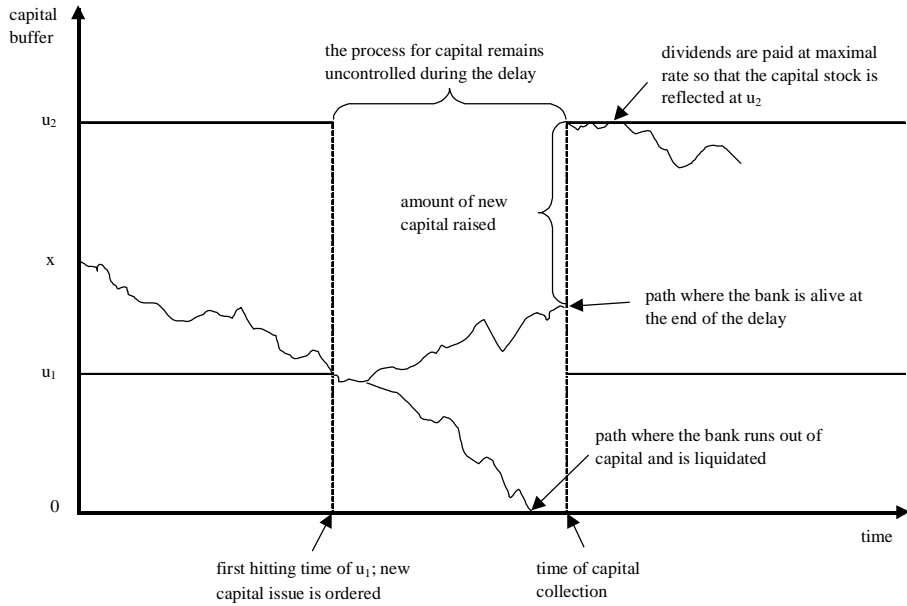


Figure 2. Illustration of the value function in a case where both u_1 and u_2 are positive

Parameter values: $\mu = 2\%$, $\sigma = 2\%$, $\rho = 1/15$, $\Delta = 0.5$ years, $K = 1\%$. The optimal barriers $u_1 = 1.80\%$ and $u_2 = 6.25\%$ are marked with vertical dotted lines. The optimal dividend barrier in the absence of the equity issue option $u_0 = 6.51\%$.

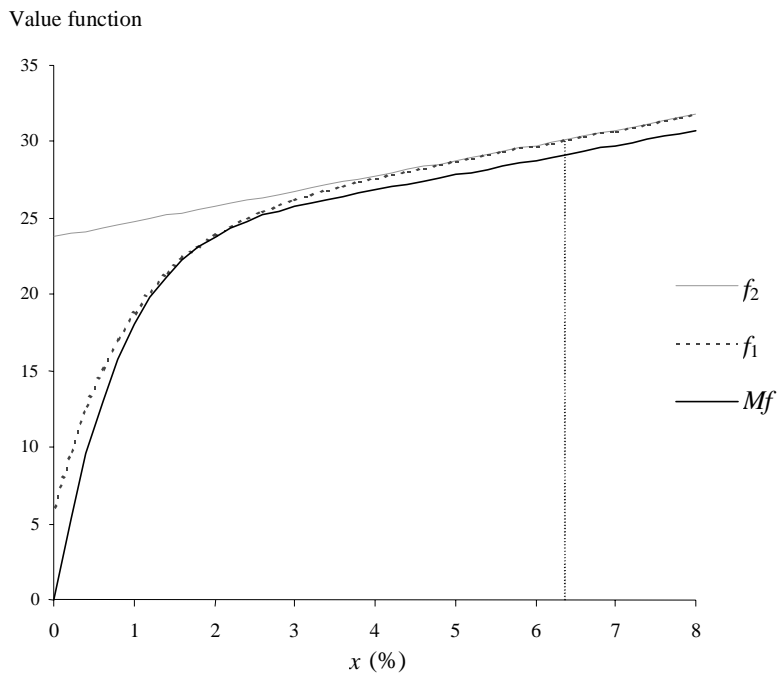


Figure 3. Illustration of the value function in a case where $u_1 = 0$

In this case $D_1M(0, u_0) < D_1f_1(0, u_0)$. Parameter values: $\mu = 2\%$, $\sigma = 2\%$, $\rho = 1/15$, $\Delta = 1$ year, $K = 6\%$. The optimal barriers $u_1 = 0$ and $u_2 = u_0 = 6.51\%$, where u_0 is the optimal dividend barrier in the absence of the option to issue new capital, given by equation (21) in the text.

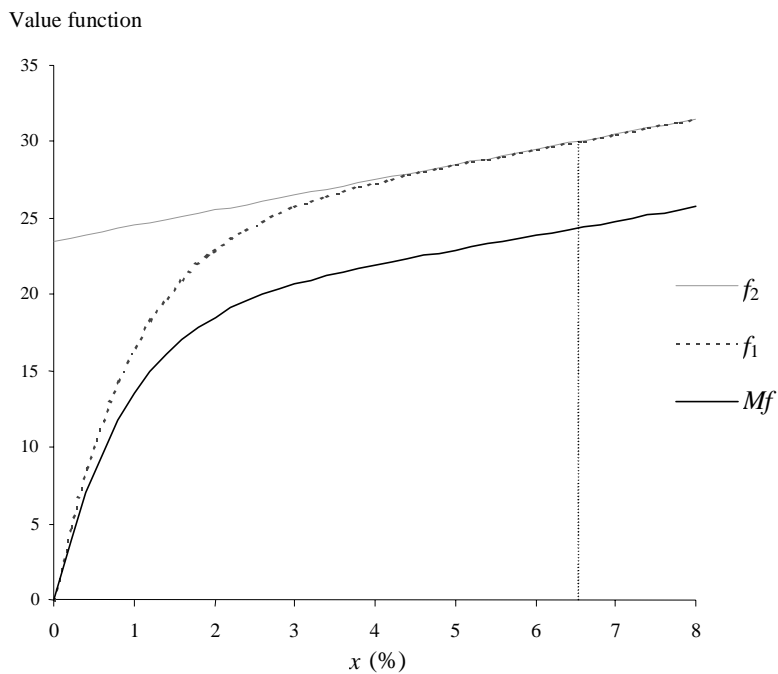


Figure 4. Model capital buffers as a function of volatility

The barrier u_2 , which is interpreted as the optimal capital ratio (in excess of the 8% minimum), as a function of volatility. The fixed parameters are equal to their calibrated values $\mu = 2\%$, $\rho = 1/15$. The fixed cost $K = 1\%$. The calibrated value for σ is 0.5%. The observed median bank capital ratio is 4.4%.

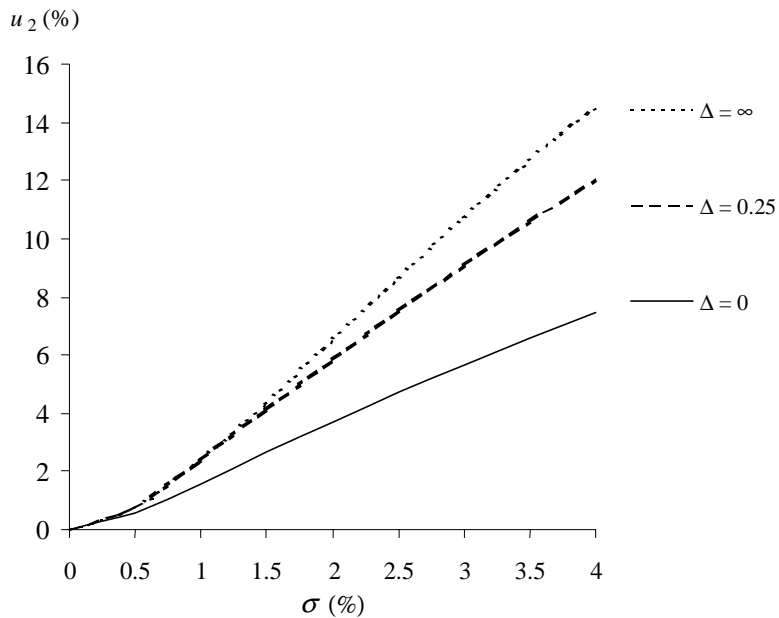


Figure 5. The dividend barrier u_2

Fixed parameter values: $\mu = 2\%$, $\sigma = 1.5\%$, $\rho = 1/15$. As Δ approaches zero, u_2 converges to its limiting value given in Proposition 3, and which is positive for positive K . As Δ or K approach infinity, u_2 converges to its maximum value u_0 , which in this example equals 4.4%.

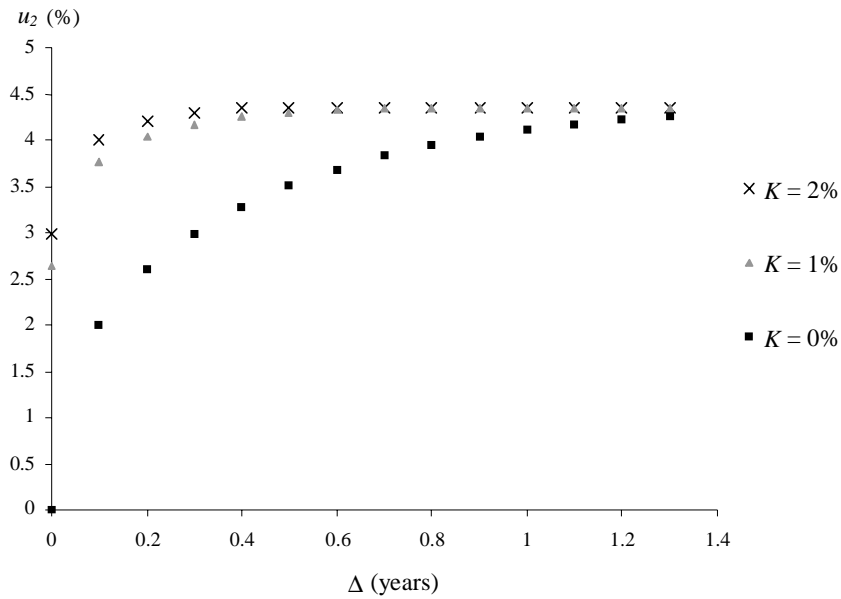


Figure 6. The capital issue barrier u_1

Fixed parameter values: $\mu = 2\%$, $\sigma = 1.5\%$, $\rho = 1/15$. As Δ approaches zero, u_1 converges to 0. As Δ or K approach infinity, u_1 converges to 0.

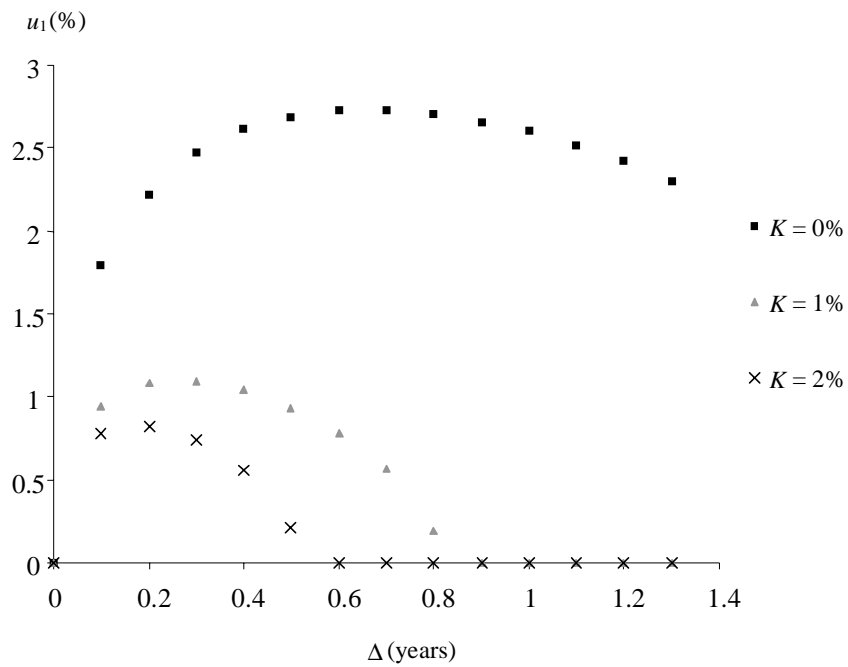


Figure 7. The value of the capital issue option for different levels of K

Fixed parameter values: $\mu = 2\%$, $\sigma = 1.5\%$, $\rho = 1/15$, $\Delta = 0.25$ years.

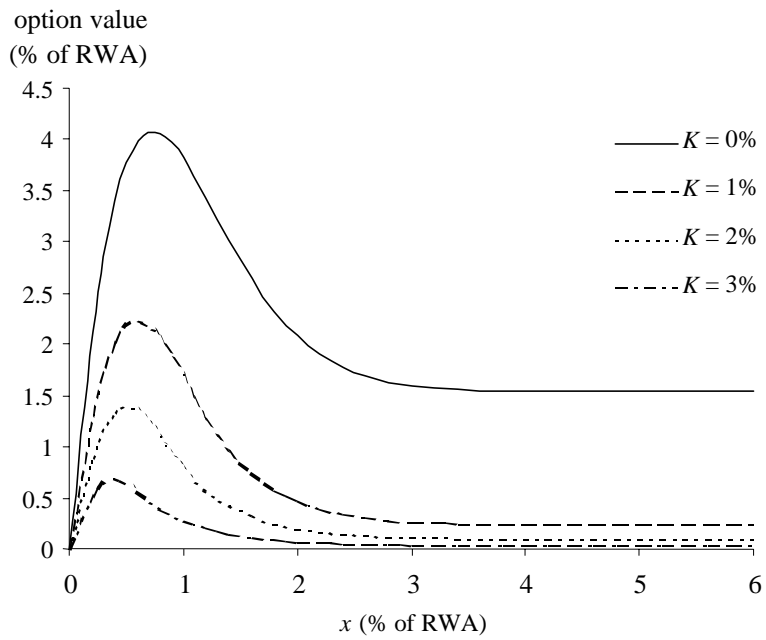


Table 1. Effective cost of capital issuance

Fixed parameter values: $\mu = 2\%$, $\sigma = 1.5\%$, $\rho = 1/15$. $E[s]$ is the expected net amount of external capital raised, calculated at the order time of a new capital issue. $K/(K+E[s])$ is the cost of capital issuance, as a proportion of the expected total amount of new capital.

Δ	K (%)	u_1 (%)	u_2 (%)	$E[s]$ (%)	$K/(K+E[s])$
0.08	0.25	1.10	3.22	1.95	11 %
0.08	0.5	1.00	3.46	2.29	18 %
0.08	1	0.89	3.70	2.64	27 %
0.08	2	0.75	3.94	3.02	40 %
0.25	0.25	1.53	3.78	1.75	13 %
0.25	0.5	1.33	3.95	2.12	19 %
0.25	1	1.09	4.11	2.52	28 %
0.25	2	0.79	4.26	2.97	40 %
0.50	0.25	1.66	4.11	1.45	15 %
0.50	0.5	1.34	4.22	1.88	21 %
0.50	1	0.93	4.30	2.37	30 %
0.50	2	0.21	4.35	3.14	39 %

Essay 3: A Value-at-Risk approach to banks' capital buffers: an application to the new Basel Accord*

Joint work with Esa Jokivuolle*

Abstract

The rating-sensitive capital charges under the New Basel Accord will increase the volatility of banks' minimum capital requirements, which may necessitate banks to hold larger capital buffers in excess of the minimum requirements. We evaluate this claim through numerical simulations on representative bank portfolios, in which bank's choice of capital buffer is assumed to satisfy a Value-at-Risk type constraint. According to our results, the size of the buffer depends on bank's credit portfolio risk and on the chosen approach to calculating the minimum capital requirement. Although the more rating-sensitive Internal Ratings Based Approach imposes lower minimum capital requirements on sufficiently high-quality credit portfolios than the Standardised Approach, this capital relief is reduced by the need for relatively higher buffers in the former approach. The buffers induced by rating-sensitive capital charges may influence banks' choice between the proposed approaches to calculating the capital requirement, as well as the overall level of bank capital after the reform. Banks' response to the volatility of the minimum capital requirements should therefore be given due consideration in the final calibration of the Basel risk weights.

Keywords: The New Basel Accord, bank capital, capital buffer, Value-at-Risk

JEL classification: G32, G35

* This is an extended version of the authors' article 'Regulatory Capital Volatility', which appeared in *Risk*, May 2001, and is printed here by the permission of *Risk Waters Ltd*.

* Esa Jokivuolle is at the Bank of Finland, P.O.Box 160, 00101 Helsinki, Finland. E-mail: esa.jokivuolle@bof.fi.

1 Introduction

The second-round proposal for The New Basel Accord in January 2001, following the first initiative published in summer 1999, is a serious attempt to increase the risk-sensitivity of the rules according to which banks are required to hold capital. The purpose of the reform is to further improve the stability of banking, to correct possible distortions in the market and to contain capital arbitrage which has gradually undermined the effectiveness of the current accord. To increase risk-sensitivity, new risk-weight categories for the currently recognized risk types are added and entirely new risk types, such as operational risks, are incorporated. Highly detailed rules are planned for taking into account various risk mitigation techniques ranging from traditional collateral to credit derivatives in determining net exposures, and treatment for securitised assets are being developed. Moreover, an overall evolutionary approach is emphasized in that the framework offers different options of varying degrees of sophistication, which banks can pursue as their resources and capabilities develop. Finally, to complement the reforms in the area of direct risk measurement and the minimum capital requirements based on them, the supervisory review process will be strengthened and banks' disclosure requirements will be increased.

Perhaps the most important and visible part of the reform is the differentiation of corporate credits, currently all receiving 100 per cent risk-weight, into a number of risk-weight categories based on credit ratings provided either by external rating agencies or banks themselves. The use of external ratings constitutes the new standardised approach (SA) in which unrated counterparties still receive the 100 per cent weight but in which rated counterparties receive weights ranging from 20% to 150%. The approach based on internal ratings (IRBA) is the more sophisticated and detailed one, also making use of additional elements such as the credit portfolio's degree of diversification. Moreover, two options, the foundation and the advanced approach, are available within the IRBA. While the SA constitutes the obligatory part of the new capital framework, a bank can choose to use the IRBA instead, subject to supervisory approval of the bank's internal credit rating systems. On the whole, the IRBA offers potentially much more risk-sensitivity than the SA. Many banks have expressed great interest in using the IRBA which would be much better aligned with

their economic capital models and which would more accurately reward low-risk portfolios with a lower capital requirement.

The increased risk-sensitivity of the regulatory risk weights will be reflected in the volatility of the regulatory capital requirement. While under the current accord shocks to banks' regulatory solvency ratios mainly result from unexpected losses, under the new framework these will also result from unexpected changes in risk-weighted assets due to ratings migration. If the volatility in minimum capital requirements rises relative to the current accord, banks are likely to increase their capital buffers above the minimum requirements. This is because – as both theory and empirical evidence since the introduction of the Basel 1988 Accord suggest - banks prefer to hold buffer capital in order to avoid the various costs, related to supervisory intervention and market discipline, which would result from approaching, let alone falling below, the regulatory minimum capital ratio (see Furfine, 2000). In this thinking the 8 per cent minimum capital ratio is analogous to a (regulatory) default point in excess of which banks need to hold their actual capital.

Banks' response to the reform will impact the overall level of capital in the banking sector. The buffers that banks prefer to hold may also have implications for banks' choice between the standardised approach and the internal ratings based approach. As the minimum capital requirement under the IRBA is likely to be substantially more volatile than under the SA, banks' capital buffers are likely to be higher under the IRBA as well (this is something that we study). The minimum capital requirement plus the required additional buffer, not the minimum capital charge alone, is likely to influence banks' choice between the two approaches. Basel obviously should take this into account in the final calibration of the risk-weights, one of the aims of which is to provide banks with capital incentives to start using the more sophisticated internal ratings based approach.

In this paper, we attempt to estimate the likely effect of the new Basel regimes on bank capital buffers. Our analysis is based on a small number of illustrative bank portfolios. We assume that a bank's choice of capital buffer satisfies a Value-at-Risk type constraint, one which states that the capital buffer should cover the bank's (credit) losses, as well as the change in the bank's minimum capital requirement, at a sufficiently high confidence level, and over a horizon which reflects the illiquidity of the bank's assets. In order to solve the capital buffer from this constraint, we implement a Monte Carlo simulation on a credit

portfolio model of the CreditMetricsTM type, but one which is extended to account for the regulatory minimum capital requirement. Effectively, we simulate bank credit losses and capital requirements simultaneously, and the Value-at-Risk requirement applies to the distribution of the sum of these two random variables. Given the complexity of the proposed Basel rules, our numerical approach is the most effective way to calculate the Value-at-Risk requirement with reasonable accuracy. As far as we are aware of, our simulation setup is novel.

Our results indicate that especially for low-risk portfolios, the capital buffers under the internal ratings based approach are likely to be substantial. Interestingly, for high-risk portfolios the additional cushions would be reduced relative to the current capital accord. A failure to account for these effects in the final risk-weight calibration could result in a different subset of banks opting for the internal ratings based approach from what is intended by Basel. We also argue that it is important to understand the way banks would likely respond to the reform when interpreting, and comparing across banks, the regulatory capital ratios in the new regime.

2 The setup

We consider a bank having all its assets in illiquid corporate loans. We argue that there are essentially three constraints that the bank would need to meet when determining how much capital to reserve against its portfolio. First, there is the minimum regulatory capital requirement associated with the bank's current portfolio. Second, the bank will want to reserve enough capital to absorb its credit losses and the fluctuations in its minimum capital requirement over its planning horizon¹, one year in our examples, with a reasonable confidence level, without recourse to new external funds. This confidence level would reflect the various implicit costs the bank would incur from approaching the minimum capital ratio. The higher these costs are, the higher a confidence level the bank would choose. A confidence level of e.g. 95% would imply that the bank would tolerate a violation of the minimum capital

¹ Our use of the term planning horizon may not be taken literally. In theory, the horizon should be related to the time it takes to liquidate the portfolio, or alternatively, to obtain external capital. Also the estimates for the key parameters needed in credit Value-at-Risk models to calculate economic capital, such as probabilities of default, are not readily available for frequencies higher than one year, which may determine the horizon to be used.

ratio on average once in every 20 years. Third, the bank will also want to satisfy an economic capital constraint, which states that capital must cover the bank's credit losses over its planning horizon, with a sufficiently high probability. This probability would typically be chosen to be consistent with the bank's overall rating target. An Aa target would imply a solvency probability of order 99.95% in a year. We suggest that the bank would choose the minimum amount of capital that satisfies each of these three constraints. We call this amount the minimum acceptable capital held by the bank. This divided by the risk-weighted assets is called the minimum acceptable capital ratio.

We develop the capital requirements here from an accounting identity governing the bank's capital dynamics. We let E_0 denote current bank capital (ex time 0 dividends), I_1 be the bank's profit, prior to any credit losses, during period 1, and L_1 be the bank's credit losses during period 1. We denote by R_t^k the bank's regulatory capital requirement at time $t \in \{0,1\}$, and under regulatory capital regime $k \in \{CA \text{ (current approach)}, SA \text{ (new standardised approach)}, IRBA \text{ (internal ratings based approach)}\}$. We assume the period length to be one year, but the formulas here apply for any period length. Finally, we define the bank's capital buffer in the capital regime k , B_t^k , as

$$B_t^k = E_t - R_t^k.$$

Assuming that no time 1 dividends are paid, the bank's capital dynamics satisfies the accounting identity

$$E_1 = E_0 + I_1 - L_1,$$

where E_1 denotes time 1 bank capital. We denote $E[L_1]$ the time 0 expectation of credit losses during year 1. We assume that the bank's pricing of its loans is actuarially fair, so that its profit before credit losses equals its expected credit losses

$$I_1 = E[L_1].$$

Credit losses over period 1 can be decomposed into expected losses and unexpected losses

$$L_1 = E[L_1] + UL_1,$$

where UL_1 is the unexpected credit loss, defined by $UL_1 = L_1 - E[L_1]$. Combining the previous three formulas, we obtain the capital dynamics

$$E_1 = E_0 - UL_1.$$

From this simple dynamics we obtain the constraints that determine bank capital buffers. The first condition is the current minimum capital requirement

$$E_0 \geq R_0^k \Leftrightarrow B_0^k \geq 0. \quad (1)$$

The second condition is concerned with capital adequacy at time 1. The bank's capital at time 1 satisfies the time 1 regulatory capital requirement if $E_0 - UL_1 \geq R_1^k$. This may be expressed in terms of the time 0 capital buffer as

$$B_0^k \geq UL_1 + \Delta R_1^k,$$

where $\Delta R_1^k = R_1^k - R_0^k$ is the change in the bank's regulatory capital charge between time 0 and time 1. Denoting the required confidence level for regulatory capital adequacy at time 1 by α , we obtain the Value-at-Risk constraint

$$P[B_0^k \geq UL_1 + \Delta R_1^k] \geq \alpha, \quad (2)$$

which is a requirement on time 0 capital buffers. It is clear from (2) that in order to calculate the minimum acceptable capital buffer, we must identify the distribution of the sum of two random variables, the unexpected losses and the change in the minimum capital requirement. Both of these are driven by rating transitions and defaults in the bank's portfolio.

The third condition for capital adequacy is what is commonly known as 'economic capital constraint', which states that the bank's time 1 capital must be non-negative, at a sufficiently high confidence level β

$$P[E_1 \geq 0] \geq \beta.$$

This may be expressed as a requirement on time 0 capital buffers as

$$P[B_0^k \geq UL_1 - R_0^k] = P[B_0^k \geq UL_1 + \Delta R_1^k - R_1^k] \geq \beta. \quad (3)$$

Because the regulatory capital charge is non-negative, a comparison of (2) and the second expression in (3) reveals that if the confidence levels α and β are equal, the constraint (2)

imposes a higher initial capital buffer. In practical applications, however, we usually have $\beta > \alpha$. In fact one of the advantages of our dynamic regulatory capital criterion (2), over the economic capital criterion (3), is that the applied confidence level need not be as close to 1, making it easier to achieve a given numerical accuracy in Monte Carlo simulations. Numerical simulations indicate that for typical values of α and β (such as 99% and 99.9%) and for average bank portfolios, the economic capital requirement (3) is not binding, and the dynamic regulatory requirement (2) determines the minimum acceptable capital buffer.

Assuming that (2) determines the minimum acceptable capital, we let \hat{B}_0^k the smallest value that satisfies (2). Then minimum acceptable bank capital is $R_0^k + \hat{B}_0^k$, and the minimum acceptable capital ratio can be expressed as

$$\left(\frac{R_0^k + \hat{B}_0^k}{R_0^k} \right) \cdot 8\% . \quad (4)$$

We explain briefly how we solve for the minimal initial capital satisfying criteria (2) and (3). We denote the distribution function of the random variable $UL_1 + \Delta R_1^k$ with F , and the distribution of the random variable UL_1 with G . Given a hypothetical portfolio, we solve for F and G from a credit portfolio model of the CreditMetrics™ type, which is appropriately modified to produce realisations of $UL_1 + \Delta R_1^k$ and UL_1 under the alternative capital regimes. In the model, latent correlated normal random variables representing customer firms' asset values are used to simulate correlated rating migrations and defaults (see J.P. Morgan, 1997, for details). In terms of F and G , we can restate the constraints (1-3) as

$$B_0^k \geq 0, \quad (1')$$

$$B_0^k \geq \min\{b \mid F(b) \geq \alpha\}, \quad (2')$$

$$B_0^k \geq \min\{b \mid G(b) \geq \beta\} - R_0^k, \quad (3')$$

so that the solution reduces to finding the percentile points of the distributions F and G corresponding to α and β , respectively.

2.1 Discussion of the set-up

Our reduced form approach is intended to be a tool for capital adequacy planning in banks. We introduce a probabilistic constraint associated with regulatory capital adequacy into a standard credit portfolio model. We simulate (changes in) regulatory capital requirements as well as unexpected losses simultaneously, and determine the required capital buffer based on their joint dynamics from a Value-at-Risk type criterion. Our extension is non-trivial because for reasonable choices for the confidence levels α and β , the dynamic regulatory capital constraint (2) is the binding one and determines the required capital buffer. Economic capital in the standard sense of the term, i.e. the solution to (3), is mostly redundant within our approach.

In a fully optimizing model of a bank, the optimal level of bank capital would be determined as a trade-off between the various benefits and costs associated with holding capital (see e.g. Furfine, 2001 or Estrella, 2001). Without doubt, a fully optimizing approach would only be feasible at the cost of abandoning the realistic model for bank portfolio dynamics considered here, as well as the detailed consideration of the Basel II rules. The confidence levels α and β in our model reflect the many tradeoffs that would be explicitly present in a fully optimizing model. In particular, the values of α and β reflect the costs and penalties associated with a violation of the regulatory capital constraint, the capital market frictions that affect the recapitalization of the bank, the sensitivity of the bank's funding cost to the amount of capital held by the bank, and the availability of growth options to the bank.

3 Numerical results

3.1 Example portfolios and parameter values

We perform our analysis under three capital regimes: the current Basel Accord (CA), the Standardized Approach under Basel II (SA), and the foundation Internal Ratings Based Approach under Basel II (IRBA). The minimum capital requirements under the three regimes

are displayed in Table 1². As for the IRBA regime, the minimum capital requirements are based on long term average annual default frequencies in Moody's data, and a uniform 50% recovery rate assumption, consistent with the proposed new Basel rules for uncollateralized senior exposures.

Table 1. Minimum capital requirements in the three capital regimes

PD = probability of default, CA = current approach, SA = standardized approach, IRBA = internal ratings based approach. The IRBA capital requirements are based on default probabilities which are from Moody's data over the period 1970-1999, and an assumed LGD of 50 %.

Rating	PD	CA	SA	IRBA
Aaa	0.01 %	8 %	1.6 %	1.1 %
Aa1	0.02 %	8 %	1.6 %	1.1 %
Aa2	0.02 %	8 %	1.6 %	1.1 %
Aa3	0.03 %	8 %	1.6 %	1.1 %
A1	0.05 %	8 %	4 %	1.5 %
A2	0.06 %	8 %	4 %	1.7 %
A3	0.09 %	8 %	4 %	2.2 %
Baa1	0.13 %	8 %	8 %	2.8 %
Baa2	0.16 %	8 %	8 %	3.1 %
Baa3	0.70 %	8 %	8 %	8.0 %
Ba1	1.25 %	8 %	8 %	11.5 %
Ba2	1.79 %	8 %	8 %	14.4 %
Ba3	3.96 %	8 %	8 %	23.2 %
B1	6.14 %	8 %	12 %	29.7 %
B2	8.31 %	8 %	12 %	35.0 %
B3	15.08 %	8 %	12 %	47.2 %
Caa	29.87 %	8 %	12 %	50.0 %

We consider nine hypothetical corporate debt portfolios, which differ in two risk dimensions: the obligor ratings distribution (aggregate, investment grade and non-investment grade), and the degree of exposure concentration (500, 100 and 50 obligors, respectively). Total exposure within each portfolio is 100, and a given portfolio always consists of loans of equal size. Each obligor is assumed to be equivalently rated externally and internally, based on the rating scale

² In the SA exposures are assigned a capital charge according to their rating only. In the IRBA capital charges are derived from a continuous function based on the obligor's probability of default and the exposure's loss given default. Moreover, an adjustment depending on the portfolio's degree of exposure concentration is added to the total capital requirement calculated in the first stage. The details of calculating the capital charge in the SA and in the IRBA are described in Basel Committee (2001). In the CA all exposures regardless of rating receive a common 8 per cent capital charge.

of Moody's³. To obtain a realistic distribution of obligors over rating categories, we use annual data over the period 1987-99 on total U.S. corporate long-term debt by rating category as rated by Moody's. In the aggregate portfolio, the share of obligors in each rating category corresponds to the time-series average of the share of that rating category on total debt over the sample period. The investment grade and the non-investment grade portfolios are constructed in a similar fashion. In the former, we only include debt in investment grade categories each year, whereas the latter includes total debt in categories Ba1 or below. All exposures in each portfolio are assumed to be senior and uncollateralized. Table 2 displays the rating distributions of the example portfolios.

Table 2. Compositions of example portfolios

Each cell gives the percentage of portfolio nominal value invested in the particular rating category. The average probability of default has been calculated using on the probabilities in Table 1. The portfolio weights are from Moody's data 1987-1999.

Rating	Investment grade	Aggregate	Non-inv. grade
Aaa	7.8 %	6.2 %	
Aa1	3.0 %	2.4 %	
Aa2	6.7 %	5.4 %	
Aa3	12.4 %	9.9 %	
A1	16.9 %	13.6 %	
A2	16.9 %	13.5 %	
A3	11.1 %	8.9 %	
Baa1	9.0 %	7.2 %	
Baa2	8.7 %	7.0 %	
Baa3	7.4 %	5.9 %	
Ba1		2.5 %	12.8 %
Ba2		2.1 %	10.9 %
Ba3		2.7 %	13.8 %
B1		3.0 %	15.2 %
B2		4.0 %	20.1 %
B3		3.8 %	19.0 %
Caa		1.6 %	8.2 %
Average PD	0.1 %	1.9 %	8.8 %

³ In calculating the IRBA capital requirements we implicitly assume that the bank is using an internal rating system that would produce equivalent ratings and default probabilities as if it were using the Moody's system and data. The Moody's system consists of 17 nondefault rating categories ranging from Aaa to Caa. According to the Basel Committee (2001), the IRBA approach would only require a minimum of 6-9 rating grades (for performing loans), so the 17 categories represents a case where the bank takes full advantage of the risk-discriminating opportunities within the IRBA. The number of rating categories used and the probabilities of default assigned to these categories will affect our quantitative results, but hardly their qualitative implications.

In simulating losses and capital requirements for these portfolios, we use a rating transition probability matrix for one-year horizon which is based on annual transition frequencies over the period 1970-1999 reported by Moody's⁴. The default probabilities from the same data are used in the calculation of the IRBA capital charges. We assume a uniform 20% asset correlation between each pair of obligors, as well as a uniform recovery rate of 50% across all obligors, corresponding to the underlying assumption made by the Basel Committee in producing the capital charges for the IRBA.

3.2 Capital requirements

Table 3 shows the statistics of the credit loss distribution and the economic capital constraints for each portfolio. In this table, the unexpected loss corresponding to a confidence level β is the β th percentile of the distribution G of the random variable UL_1 . The unexpected loss with confidence level β hence coincides with the economic capital requirement under the same

Table 3. Credit risk statistics and the economic capital constraint

Unexpected loss is defined as the given percentile of the credit loss distribution less expected loss. Total exposure in each portfolio equals 100.

Portfolio		Expected loss	Standard deviation	Unexpected loss		
				$\beta = 99\%$	$\beta = 99.9\%$	$\beta = 99.95\%$
Investment grade	“500”	0.1	0.1	0.6	1.5	1.8
	“100”	0.1	0.2	1.0	1.9	2.4
	“50”	0.1	0.3	1.5	2.1	3.0
Aggregate	“500”	1.0	0.8	2.8	4.8	5.8
	“100”	1.0	1.0	3.5	5.5	6.5
	“50”	1.0	1.2	4.0	6.0	7.0
Non-investment grade	“500”	4.4	3.5	11.8	18.4	20.1
	“100”	4.4	3.7	12.1	19.1	21.1
	“50”	4.4	3.9	13.5	19.6	22.1

⁴ Again, these are assumed to be representative of the probabilities with which the model bank updates its internal ratings. We note that banks may use rating systems with different quantitative transition characteristics, which entail higher off-diagonal transition probabilities than the Moody's matrix (see e.g. Kealhofer et al., 1998). This would further increase the volatility in minimum capital requirements.

confidence level, expressed as a minimum level of equity (rather than as a capital buffer). We observe from Table 3, as expected, that the degree of diversification of the portfolio does not influence expected losses, whereas measures of loss dispersion increase as diversification is reduced. For a well-diversified portfolio with 500 assets, the standard deviation of the credit loss distribution is of the same order of magnitude than the mean of the distribution. We observe that diversification still offers considerable benefits for credit portfolios with over 100 assets. In relative terms, the benefits of diversification appear the higher, the better the portfolio credit quality. Finally, Table 3 provides evidence that the skewness of the credit loss distribution is the more extreme, the better the average credit quality of the portfolio. For an investment grade portfolio with 500 assets, the ratio of the 99.9th percentile of the distribution to the standard deviation of the distribution is 15. For a corresponding non-investment grade portfolio the same ratio is only around 5.

Table 4. Minimum capital requirements

The regulatory 8 per cent minimum capital requirements. Total exposure in each portfolio equals 100. The lower of the minimum capital requirements in the SA and in the IRBA is indicated with gray shading.

Portfolio		CA	SA	IRBA
Investment grade	“500”	8.0	4.3	2.4
	“100”	8.0	4.3	3.4
	“50”	8.0	4.3	4.7
Aggregate	“500”	8.0	5.6	7.9
	“100”	8.0	5.6	9.4
	“50”	8.0	5.6	10.8
Non-investment grade	“500”	8.0	10.5	29.9
	“100”	8.0	10.5	31.1
	“50”	8.0	10.5	33.2

In Table 4 we report the 8 per cent regulatory minimum capital requirements. We make some obvious remarks. First, while the current Basel regime does not differentiate corporate exposures according to their credit risk, the SA and the IRBA indeed do so. Second, only the IRBA punishes or rewards portfolios according to their degree of diversification, measured by exposure concentration. The effects of diversification are captured by the granularity adjustment term in the IRBA, not the baseline risk weights themselves. Table 5 shows this, by decomposing the IRBA minimum capital requirements in Table 4 into components due to the

baseline risk weights and the granularity adjustment, respectively. Table 5 further shows that the effect of the granularity adjustment is the largest in relative terms in high-quality portfolios. Moreover, both negative and positive granularity adjustments occur among our example portfolios. Third, the SA looks very competitive against the IRBA based on the minimum capital requirements. In fact the IRBA minimum capital requirement is lower than the SA minimum capital requirement in case of the 500 and 100 asset investment grade portfolios only.

Table 5. The effects of granularity adjustment in the IRBA

RWA = risk-weighted assets, GA = granularity adjustment.

Portfolio		RWA	GA	RWA+GA	8%* RWA	8%* (RWA+GA)
Investment grade	“500”	28.0	2.0	30.1	2.2	2.4
	“100”	27.7	14.7	42.4	2.2	3.4
	“50”	28.6	30.5	59.1	2.3	4.7
Aggregate	“500”	99.4	-0.5	98.9	7.9	7.9
	“100”	103.7	13.4	117.1	8.3	9.4
	“50”	103.4	31.1	134.5	8.3	10.8
Non-investment grade	“500”	385.6	-11.5	374.0	30.8	29.9
	“100”	384.4	4.1	388.6	30.8	31.1
	“50”	390.9	23.5	414.4	31.3	33.2

The minimum acceptable capital amounts, calculated from (2’), are presented in Table 6. All results are based on the confidence level $\alpha = 99\%$. In those cells in Table 6 where the economic capital constraint (3’) is binding, the resulting minimum acceptable capital amount is reported in parentheses in the respective cell. In all cases economic capital has been calculated based on a confidence level $\beta = 99.95\%$, so that the figures correspond to the right-most column in Table 3. We observe that the economic capital constraint is binding under the CA and the SA for the non-investment grade portfolios. This reflects the fact that the CA and the SA are not sensitive enough to portfolio risk, so that the bank’s own economic capital requirement comes into play. Under the IRBA, the economic capital requirement is not binding for any of the portfolios considered. This is not very surprising, knowing that a multiplier greater than one has been used in deriving the current IRBA benchmark risk-weights relative to an underlying economic capital model (for an explanation of the derivation of the IRBA risk-weights, see Wilde, 2001).

Table 6. Minimum acceptable capital amounts

The minimum amounts of capital satisfying (2) for the example portfolios, given $\alpha = 99\%$. Figures in parentheses are the economic capital requirements (3) where it is the binding constraint, given a confidence level $\beta = 99.95\%$. Total exposure in each portfolio equals 100. The lower of the capital requirements in the SA and in the IRBA is indicated with gray shading.

Portfolio		CA	SA	IRBA
Investment grade	“500”	8.5	5.6	4.8
	“100”	8.8	5.8	6.1
	“50”	9.3	6.0	7.7
Aggregate	“500”	10.3	8.4	10.6
	“100”	10.8	8.8	12.1
	“50”	11.2	9.1	13.6
Non-investment grade	“500”	17.2 (20.1)	19.0 (20.1)	33.2
	“100”	17.5 (21.1)	19.4 (21.1)	34.6
	“50”	18.7 (22.1)	20.1 (22.1)	36.6

Since the nominal value of each portfolio has been normalized to 100, the capital amounts in Table 6 also represent effective equity ratios for the portfolios. These equity ratios, which are obtained by dividing capital by nominal assets, are different from the usual bank capital ratios, which refer to capital over risk-weighted assets. The equity ratios are very informative in comparisons that take place between capital regimes, as we do here, because the risk-weighted assets used to calculate standard capital ratios are not constant across capital regimes.

Figure 1 illustrates the contents of tables 4 and 6, by decomposing the acceptable capital amounts for the 500 asset portfolios into the minimum requirement and the buffer, in each capital regime. This brings out the relative size of the buffer with respect to the minimum requirement. We note that buffers are very large in relative terms for investment grade portfolios in the IRBA, although the total acceptable capital amount in this case is quite low. Buffers are also large in relative terms for non-investment grade portfolio under the current Basel regime as well as under the SA. Perhaps surprisingly, however, the relative size of the buffer is quite low for non-investment grade portfolios in the IRBA. We provide an explanation to this below.

Figure 1. Decomposition of acceptable capital into minimum capital and capital buffer

Based on the results for the 500 asset portfolios. The figure combines information from tables 4 and 6. To focus on the effect of rating sensitivity, the economic capital constraint (the figures in parentheses in Table 6) is ignored.

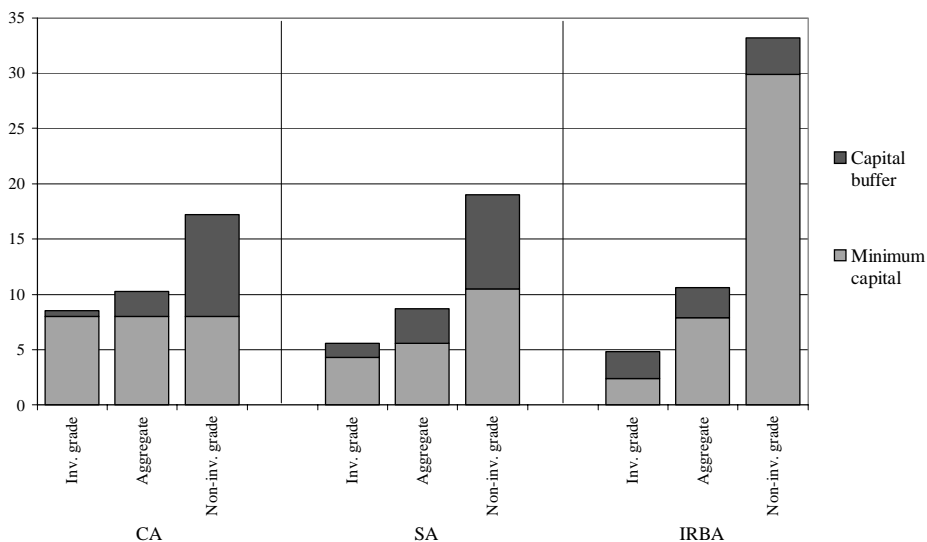


Table 7 shows the capital ratios for all the example portfolios. These have been calculated from the minimum acceptable capital amounts through formula (4). All the capital ratios exceed the 8% regulatory minimum. We observe that the highest capital ratios are for well diversified investment grade portfolios under the IRBA, and for poorly diversified non-investment grade portfolios under the current regime. In general, we observe that capital ratios of investment grade portfolio rise, going from the current regime to the more risk-sensitive regimes, while the capital ratios for non-investment grade portfolio are reduced. The capital ratios for average quality portfolios are virtually equivalent under the current regime and the IRBA. As an example, we note that the minimum acceptable capital ratio for the 500 asset investment grade portfolio in the current regime is 8.5%, whereas in the IRBA the ratio is 16.0%. The corresponding ratios for the 500 asset non-investment grade portfolio are 17.2% and 8.9%. Table 7 also shows that the capital ratios under the current regime and under the SA decline when portfolio diversification increases, whereas in the IRBA the reverse

happens. The odd behavior of the capital ratio in the IRBA with respect to portfolio diversification is due to the granularity adjustment, which results in lower minimum capital charges for well-diversified portfolios, but as Table 7 confirms, part of this capital savings is consumed by the need for relatively larger capital buffers.

Table 7. Minimum acceptable capital ratios

The minimum acceptable capital ratios for the example portfolios, given $\alpha = 99\%$, calculated from (4). Figures in parentheses indicate the minimum acceptable capital ratio in cases where the economic capital constraint (3) is binding for the most conservative choice of $\beta = 99.95\%$. Total exposure in each portfolio equals 100.

Portfolio		CA	SA	IRBA
Investment grade	“500”	8.5%	10.4%	16.0%
	“100”	8.8%	10.7%	14.4%
	“50”	9.3%	11.2%	13.1%
Aggregate	“500”	10.3%	12.0%	10.7%
	“100”	10.8%	12.7%	10.4%
	“50”	11.2%	13.0%	10.1%
Non-investment grade	“500”	17.2% (20.1%)	14.5% (15.3%)	8.9%
	“100”	17.5% (21.1%)	14.8% (16.1%)	8.9%
	“50”	18.7% (22.1%)	15.3% (16.8%)	8.8%

That the relative size of the capital buffer for a given portfolio is not always higher in the more risk-sensitive IRBA, relative to the SA or the current regime, may appear as somewhat surprising. In fact, Table 7 indicates that the IRBA gives rise to higher minimum acceptable capital ratios than the SA only for the investment grade portfolios. A related observation is that in the CA and in the SA, the minimum acceptable capital ratio increases as the risk of the portfolio increases, whereas in the IRBA the reverse happens⁵. In order to understand why, we analyze the two random components that determine the capital requirement (2). The first component is the unexpected loss, and the second component is the change in the minimum capital requirement. In a high-quality portfolio there can be mainly rating downgrades but very few losses. This implies that in a rating-sensitive capital regime, like the IRBA, rating migration can cause considerable variation in the minimum capital requirement, whereas in a

capital regime relatively insensitive to ratings this is not the case. Therefore the minimum acceptable capital ratio for the investment grade portfolio is the highest in the IRBA, the second-highest in the SA, and the lowest in the current regime, CA. As for low credit quality portfolios, the unexpected loss component now contributes significantly to the capital requirement. Hence the minimum acceptable capital ratios in the risk-insensitive capital regimes are relatively high. In the risk-sensitive IRBA, on the other hand, the opposite happens because of two reasons. First, relative to a high-quality portfolio, there is more scope for rating upgrades which the risk-sensitive regime rewards with a lower minimum capital requirement. Secondly, and perhaps more importantly, the initial capital requirement in the more risk-sensitive approach already accounts for much of the potential losses in the planning period because the lowest rated assets receive capital charges which are of the same order of magnitude as realized losses in the event of default. Indeed, the highest-risk exposures receive the cap risk-weight in the IRBA, which implies that their capital charge equals their loss-given-default. When losses are realized on such assets, there is an exactly compensating reduction in the amount of risk-weighted assets, and the net effect on the bank's capital buffer is zero.

Finally, returning to Table 6, we have indicated by gray shading the capital regime that the bank would choose on the basis of the lowest acceptable amount of capital, ignoring other factors that could affect its choice. We observe that only for the 500 asset investment grade portfolio would the bank opt for the IRBA, otherwise it would select the SA. This is a rather striking result, given that the Basel Committee and the EU have indicated that they would like to see a fairly large number of banks moving to the IRBA⁶. Moreover, it is interesting to note that only in investment grade portfolios would better diversification provide sufficient incentives to move from the SA to the IRBA.

⁵ In the IRBA this is the case in moving to portfolios of poorer credit quality, keeping the degree of diversification fixed, and also in moving to less diversified portfolios, keeping the portfolio's overall credit quality fixed. The capital ratio of non-investment grade portfolios, however, is practically insensitive to the degree of portfolio diversification.

⁶ We acknowledge that the assumption that all counterparties in our portfolios possess external ratings is likely to bias this comparison in favor of the SA. In the absence of external ratings, the SA basically collapses to the CA, which looks significantly less advantageous relative to the IRBA. Another possible reason for obtaining results strongly in favor of the SA is that we do not consider all the risk mitigation techniques which would benefit banks more in the IRBA than in the SA.

4 Implications and conclusions

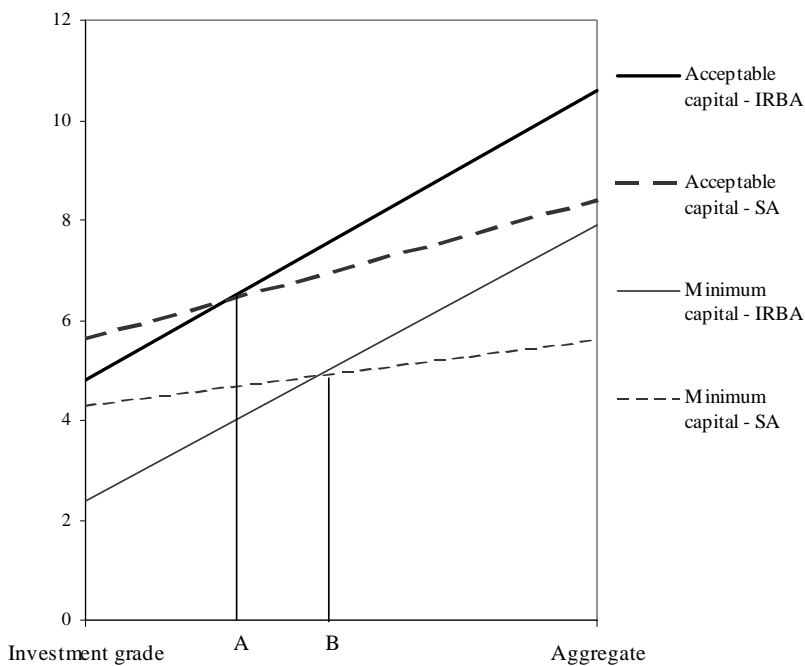
The effect of the capital regulation reform on overall bank capital depends on the amount of buffer capital that banks prefer to hold in excess of the new minimum requirements. The Basel Committee has stated as its goal that, after the reform, the overall level of capital should stay at about the same level as prior to the reform. An important implication of our analysis is that in order to achieve this goal, the banks' likely adjustment of their capital buffers should be taken into consideration in the calibration of the new capital requirements. In particular, assume that regulators' implicit criterion in determining the level of bank capital is to achieve a socially desirable level of default probabilities for banks. If the minimum capital requirement is set at a level that implies a given default probability, then the minimum acceptable amount of capital chosen by a prudent bank, in light of our results, would imply a substantially lower effective probability of default. Comparing tables 4 and 6, we observe that for investment grade portfolios the minimum acceptable capital amount under the IRBA is roughly twice the minimum requirement, indicating that actual bank default probability may turn out to be much lower than the one pursued. Banks should be made safe, but not too safe, because holding of idle capital has a cost from both microeconomic and macroeconomic points of view.

Second, the additional buffers may give certain banks a disincentive to start using the IRBA. Other things being equal, a bank would choose the regulatory approach that would give it the lowest acceptable capital level. Basel has indicated that in the final stage of the reform the IRBA risk-weights would be calibrated in such a way that a representative bank – and obviously banks with risks that are lower than those of the representative bank - would have an incentive in the form of a lower *minimum* capital requirement to use the IRBA rather than the SA. A calibration done in this way could lead to problems such as in the case of the investment-grade 100 asset portfolio, for which the minimum capital requirement in the IRBA is lower than in the SA, whereas the minimum acceptable capital level in the SA is lower than in the IRBA (see the different coverage of the shadings in tables 4 and 6). This incentive problem is further illustrated in Figure 2. There we have graphed out the simulation results for the 500 asset portfolios from Tables 4 and 6. These show that the set of portfolios for which the minimum acceptable capital in the IRBA is lower than in the SA (portfolios to the left of A), is different than the set of portfolios for which the minimum capital requirement in the

IRBA is less than in the SA (portfolios to the left of B). Hence ignoring the capital buffers could lead Basel to provide capital incentives towards using the IRBA for a somewhat smaller set of banks from what was intended.

Figure 2. Illustration of capital incentives

The figure illustrates the fact that because of its lower risk-sensitivity, the SA may be a preferred alternative to the IRBA, when evaluated based on minimum acceptable capital amounts (which take into account the required capital buffer), even though the IRBA yields a lower minimum capital requirement (this happens between points A and B). Based on numerical results for the 500 asset portfolios in tables 4 and 6.



Third, we suggest that after the reform, banks’ capital ratios will have to be interpreted with care. In the IRBA, the lowest-risk banks are likely to have the highest capital ratios (i.e. capital relative to risk weighted assets), although their absolute capital amounts will be the lowest. Consequently, unlike in the current accord or in the SA, a high capital ratio of an IRBA bank will not be an indication of high portfolio risk, but to the contrary. We suggest that a simple leverage ratio – capital over nominal, not risk-weighted, assets - is more directly related to a bank’s portfolio risk after the reform, and that banks operating in different capital

regimes can only be compared based on common denominator. In our example the portfolio size has been normalized to 100, so that the minimum acceptable capital amounts in Table 6 can be directly interpreted as such leverage indicators.

Fourth, our analysis indicates that banks may have incentives to try to dampen the volatility in their capital requirements, particularly so in the IRBA where banks are allowed to use their internal ratings. This may give rise to new types of capital arbitrage activities, where banks obtain the image benefits from being an 'IRBA bank', while simultaneously attempting to smoothen the variability in their internal ratings, so as to keep regulatory capital volatility at a low level.

Finally, our analysis suggests that banks' adjusting their capital cushions may help in alleviating the pro-cyclical effects of more risk-sensitive capital requirements that have been pointed out in the literature (see Blum and Hellwig, 1995). An increase in the volatility of banks' minimum capital requirements, *ceteris paribus*, will increase the probability that a bank will face a regulatory capital shortage in an economic downturn. A regulatory capital shortage, on the other hand, may force banks to cut back on lending or to liquidate assets, and therefore provides an enforcing feedback to the original shock in the economy. Increased capital buffers will help banks to tolerate larger fluctuations in their minimum capital requirements, alleviating the effects of this procyclical mechanism.

References

Basel Committee for Banking Supervision, 2001, The New Basel Capital Accord, BIS, January.

Blum, J., M. Hellwig, 1995, The macroeconomic implications of capital adequacy requirements for banks, *European Economic Review*, 39, 739-749.

Estrella, A., 2001, The cyclical behavior of optimal bank capital, working paper, Federal Reserve Bank of New York, New York.

European Commission, 2001, Commission Services' Second Consultative Document on Review of Regulatory Capital for Credit Institutions and Investment Firms, February 2001.

Furfine, C., 2000, Evidence on the response of US banks to changes in capital requirements, BIS Working Paper, June.

Furfine, C., 2001, Bank portfolio allocation: the impact of capital requirements, regulatory monitoring, and economic conditions, *Journal of Financial Services Research*, 20, 1, 33-56.

J.P. Morgan, 1997, *CreditMetricsTM - Technical Document*, New York.

Jokivuolle, E., S. Peura, 2001, Regulatory capital volatility, *Risk*, 14, May, 95-98.

Kealhofer, S., S. Kwok and W. Weng, 1998, Uses and abuses of bond default rates, *J.P. Morgan CreditMetrics Monitor*, First Quarter, 37-55.

Kupiec, P. H., 1999, Risk capital and VaR, *Journal of Derivatives*, 7, 2.

Wilde, Tom, 2001, IRB Approach Explained, *Risk*, 14, May, 87-90.

Essay 4: A structural model of risky debt with stochastic collateral*

Joint work with Esa Jokivuolle*

Abstract

We present an extension of the Merton model of risky debt in which collateral value is a separate random variable correlated with the probability of default. The model is particularly suited for studying the behaviour of the expected loss-given-default, as a function of collateral value parameters, and could be used for estimating losses-given-default in many popular models of credit risk which assume them constant. We also examine the problem of determining sufficient collateral amount to secure a loan to a desired degree. The estimation of the expected loss-given-default is of increased interest to bank practitioners and regulators due to the proposed new Basel Accord.

Keywords: debt valuation, collateral, loss-given-default, loan-to-value

JEL classification: G13, G21

1 Introduction

The effect of collateral values, and recovery rates in general, on the value of defaultable debt is evident from the generic debt valuation formula

$$F = PV(1 - ELGD \cdot PD). \quad (1)$$

The value F of a risky dollar to be received in the future is obtained by applying an appropriate present value operator to the expected payoff at maturity, where the expected payoff is calculated by deducting the product of expected losses-given-default (ELGD) and

* This is a modified version of the authors' paper 'Incorporating collateral value uncertainty in loss-given-default estimates and loan-to-value ratios', forthcoming in *European Financial Management*, 2003, and is printed here by the permission of *Blackwell Publishing Ltd*.

* Esa Jokivuolle is at the Bank of Finland, P.O.Box 160, 00101 Helsinki, Finland. E-mail: esa.jokivuolle@bof.fi.

the probability of default (PD), from the promised payment. The formula implies that a 10% change in the ELGD has an equivalent effect on the value of risky debt as a 10% change in the probability of default. In this sense, the ELGD and the probability of default are equally important determinants of risky debt values. Moreover, collateral value, the main determinant of ELGD, is in many cases an observable and traded quantity, whereas default probabilities are not directly observable. These arguments suggest that the modelling of collateral value dynamics should be of high priority in the valuation of risky debt.

In the seminal work on debt valuation based on structural models of Merton (1974), the asset value (which determines default) and the collateral value (which determines the payoff to risky debt in the event of default) are the same process. Consequently, the correlation between the default determining variable and the collateral value in this model is perfect. It should be of interest to analyse situations where the determinant of default and the value of collateral are less than perfectly correlated. An example of such a situation is a bank loan backed by a pledged asset which is not owned by the borrowing firm but by a third party. In this case it is possible that the collateral value is relatively high even though the borrower's asset value is low, or vice versa. In general, however, we would expect the collateral value and the borrower's default probability to be negatively correlated. The values of most assets depend positively on overall business conditions, so that in a bad macroeconomic realisation where most defaults take place, the values of collateral items are likely to be low as well¹.

In this paper, we develop and analyse a model of risky debt where the default probability of the borrower and the value of the collateral supporting the debt obligation are less than perfectly correlated. Our model extends Merton's (1974) structural model in that we add a separate collateral value process which is correlated with the default determining asset value process. We interpret the collateral variable as the market value of the assets that the debt holder has a claim on in the event of a liquidation sale. We do not attach any particular asset stock interpretation to the underlying asset process, but rather think of the normalised difference between the asset value and the default boundary, representing a measure of distance to default, as a sufficient statistic for the likelihood of default. For this reason, we use

¹ Schleifer and Vishny (1992) present an equilibrium analysis of why collateral values tend to decline just when defaults abound.

the term *default determining variable* interchangeably with asset value. In line with this interpretation, our preferred implementation of the model does not require estimation of the asset value parameters individually, but relies on a one-to-one mapping between the distance to default measure and the default probability. Our non-concrete interpretation of the underlying asset value is similar to the interpretation suggested e.g. by Longstaff and Schwartz (1995)². Alternatively, since our structural assumptions restrict the evolution of default probabilities, but without implying a stringent interpretation of the default determining variable, our model can be classified as a reduced form model with special structure. The classification into structural and reduced form models of debt valuation is a shady one, and we think that our model can be interpreted to represent either class.

Our model is particularly suited for answering two questions of great interest to bank practitioners as well as to regulators. The first is the issue of how should losses given default, that are stochastic, be estimated for the purpose of using them in credit risk models that often for practical reasons take them as constants. We provide a simple expression for the ELGD within our model, which could be the basis for loss-given-default estimates to be used in several reduced form models of risky debt (e.g. Jarrow and Turnbull, 1995, Jarrow, Lando and Turnbull, 1997, and Duffie and Singleton, 1999), as well as in portfolio models such as J.P. Morgan (1997) or Credit Suisse Financial Products (1997). In particular, when credit portfolio models are applied to bank loan books with a very large number of individual exposures, loss-given-defaults are often treated as constants for computational reasons. The question is then to find the constant estimates that give the best approximation to the results obtained from a full-scale portfolio model with stochastic collateral values. The ELGD for each individual exposure, calculated on a stand-alone basis, is a natural, yet not necessarily the theoretically correct, estimate to this end³. Effectively the same problem is also encountered by regulators who are currently working on reforming capital adequacy

² Longstaff and Schwartz (1995) assume an asset value process identical to ours, and a constant boundary whose first-crossing time determines default. They offer both an asset stock and a cash flow based interpretation to this default mechanism. Moreover, they emphasize that the critical determinant of default is the value of the default determining variable relative to the default boundary, and not the value of the default determining variable in itself.

³ In portfolio models, conditional expectations are frequently substituted for the values of stochastic variables in order to speed up computations. The correct conditioning event, however, depends e.g. on the chosen Value-at-Risk confidence level. See e.g. Praschnik, Hayt, and Principato (2001) on this.

requirements to be more in line with economic capital models (Basel Committee, 2001). The internal ratings based approach to setting capital requirements on credit risk within the new Basel Accord is based on a one-factor version of the CreditMetricsTM framework, and the issue of how to estimate loss-given-defaults consistently with the dynamics of collateral values is also present here.

As a second application, we suggest that our model could be used by banks in setting limits on loan-to-value ratios to be applied in lending. A limit on the loan-to-value ratio refers to the maximum amount of credit that may be granted against a given collateral. Such limits appear to be widely used by banks, and may partially substitute for risk sensitive pricing of loans. Often, however, there may not be a transparent quantitative theory which could explain the structure of the chosen limit system. We believe that existing practices could potentially be much improved upon by a model based quantitative approach. To this end, we calculate limits on loan-to-value ratios based on an uniform value principle, in which the choice of the loan-to-value ratio equalizes the valuation of loans across risk classes, as well as two other related probabilistic criteria.

Our model yields a number of comparative static predictions on the behaviour of the ELGD with respect to the model parameters. We show that the ELGD is a decreasing function of the drift of the collateral value, an increasing function of the volatility of the collateral value, and an increasing function of the correlation between the collateral value and the default determining variable. Moreover, the ELGD is a decreasing function of the initial default probability of the borrower, given that the correlation between the collateral and the default probability is negative (the usual case). This last result at first appears counter-intuitive, but we present an explanation to it. Our numerical results indicate that this effect may not be negligible in all cases.

There is a large literature on the analysis of risky debt based on structural models. The seminal model of Merton (1974) has been extended to allow for endogenous bankruptcy e.g. by Black and Cox (1976), Leland (1994), and Leland and Toft (1996). Stochastic interest rates have been considered by Kim, Ramaswamy and Sundaresan (1993), Longstaff and Schwartz (1995), and Briys and de Varenne (1997). Liquidation costs and the resulting bargaining game in the event of bankruptcy have been analysed by Andersen and Sundaresan

(1996), Fan and Sundaresan (2000), and Mella-Barral and Perraudin (1997). However, the only analysis that has considered the effects of less than perfect correlation between the firm asset value and the collateral value, that we are aware of, is Frye (2000a, 2000b). Frye uses the ELGD concept in proposing improved constant loss-given-default estimates in a portfolio context, where asset values and collateral values are assumed to be normally distributed. Our qualitative results regarding the ELGD are consistent with his, but we believe that the Mertonian lognormal framework that we use is a more realistic model of asset and collateral values.

The rest of the paper is organized as follows. The next section presents our extension of the Merton model. Section 3 solves the model for debt prices and derives some of their properties which are relevant concerning the implementability of the model. Section 4 analyses the comparative statics of debt values and the ELGD, while Section 5 illustrates loan-to-value ratios derived from the model. Section 6 contains a discussion of the potential applications of our model within the new Basel Accord.

2 The model

We build on Merton's classic model of risky debt. We study a defaultable zero-coupon debt contract with a face value of B , and a maturity of T years, which is backed by stochastic collateral V . The following assumptions characterize our model.

Assumption 1. *The default determining variable A satisfies the stochastic differential equation*

$$dA = \mu_A A dt + \sigma_A A dW_1, \quad (2)$$

where the drift μ_A and the diffusion term σ_A are positive constants, and $\{W_{1t} : t \geq 0\}$ is a standard Brownian motion.

Assumption 2. *The collateral value V satisfies the stochastic differential equation*

$$dV = \mu_V V dt + \sigma_V V dW_2, \quad (3)$$

where μ_V and σ_V are positive constants, and $\{W_{2t} : t \geq 0\}$ is a standard Brownian motion, whose instantaneous correlation with W_1 is ρdt .

We interpret V as the market value that the lender can realize in an asset sale, when she liquidates both the collateral explicitly pledged against the debt obligation, as well as her share of the nonpledged assets. Thus we use the word collateral in a wide sense, and not only to mean the value of the pledged assets.

Assumption 3. *The risk-free interest rate r is independent of the variables A and V .*

Since our objective is to study the behaviour of credit spreads and loss-given-defaults, we can assume without loss of generality that the riskfree rate is a constant. We do this in all subsequent presentation to economize on notation.

Assumption 4. *Default occurs at T when the value of the default determining variable A at T is less than a positive constant D .*

According to the classical interpretation of A as total asset value, default is triggered if the asset value at maturity is less than the face value of the zero coupon debt. In this case our default boundary D would equal the face value of the zero coupon debt, B . We do not need to make such stringent interpretations concerning the capital structure of the firm. Even if we interpreted A as total asset value, we could allow several forms of debt by defining D as the total amount of debt. As suggested in the introduction, we prefer not to think of A as total asset value, but rather think of A and D jointly as a mechanism that generates a term structure of default probabilities for the firm. This interpretation is supported by the fact (which will be shown in the next section) that debt value is homogenous of degree zero in A and D . Thus only their relative magnitude matters for pricing of credit risk, not their absolute values.

Default only occurs at the maturity of the debt contract in our model. We acknowledge that this is a shortcoming of the Merton model, and precludes coupon debt from being priced consistently. The portfolio models, on the other hand, which are the intended application for the ELGD estimates calculated from our model, share a similar one-periodic structure to our model.

Assumption 5. *The payoff to the holder of the zero coupon debt at T equals*

$$BI(A_T \geq D) + \text{Min}(V_T, B)I(A_T < D), \quad (4)$$

where $I(\cdot)$ denotes the indicator function for the event defined in the parenthesis.

Assumption 5 implies that the payoff to the debt holder at maturity depends on A and V through their terminal values only. The value of A alone determines whether the firm is in default at T or not. If the value of A at T exceeds the default barrier D , the debt holder receives the promised amount B . When the value of A at maturity is below D , the firm is in default and the holder of the zero coupon debt receives $\text{Min}(V_T, B)$. This implies that a debt holder can never benefit from default through receiving more than her promised share of the value of the firm.

So far we have assumed that A and V are correlated but have not defined their functional relationship. According to the asset value interpretation of A , there are basically two alternatives. If the debtor owns the collateral, then V is part of A . It is also possible that the collateral is not owned by the debtor but is provided by a third party. A common example of this would be an entrepreneur whose privately held company is the debtor but who provides the collateral to the lender as a private person. In this case V and A are correlated inasmuch as macroeconomic developments tend to affect all asset values in the economy, whether they be corporate assets or e.g. housing real estate.

In the case where the debtor owns the collateral, we could specify $A = A' + V$, where A' and V are correlated lognormal diffusion processes. The correlation between A and V would then be effectively derived within the model from the processes of A' and V . As the sum of two lognormal variables is not lognormal, such additive relation between A and V is not likely to work analytically. We have chosen to keep V and A separate, which, if the asset value interpretation is taken literally, corresponds to the case where the debtor does not own the collateral. But since we do not enforce the asset value interpretation in the first place, there is no need to combine A and V in any way. After all, if A is only a sufficient statistic for a default probability, A and V do not need to be in the same scale, and trying to force any summing between them would be a pointless exercise.

Assumption 6. *Sufficiently perfect capital markets.*

In our analysis, we ignore transactions costs, taxes and indivisibilities, and assume that agents are price takers. Short sales are allowed, and trading can take place continuously in time. Many of these assumptions can be relaxed without changing the conclusions materially.

3 Value of risky debt

Let $F(A, V, t)$ denote the value of the zero coupon debt as a function of the stochastic variables and (calendar) time t . At maturity the value is just the terminal payoff to the zero coupon debt, given by equation (4) above

$$F(A_T, V_T, T) = BI(A_T \geq D) + \text{Min}(V_T, B)I(A_T < D).$$

Simple manipulations yield an equivalent form of the terminal payoff

$$F(A_T, V_T, T) = B - \text{Max}(0, B - V_T)I(A_T < D), \quad (5)$$

which reveals that the payoff from this risky debt is equivalent to a payoff from the corresponding riskfree debt, less the payoff from a European put option written on the collateral value, with strike B and maturity T , whose realisation is conditional on the value of the stochastic variable A . We will utilize this representation in what follows.

Time 0 expected payoff, under an arbitrary probability P , is

$$\begin{aligned} E^P[F(A_T, V_T, T)] &= B - E^P[\text{Max}(0, B - V_T)I(A_T < D)] \\ &= B - E^P[\text{Max}(0, B - V_T)I(A_T < D)]E^P[I(A_T < D)] \end{aligned} \quad (6)$$

We define formally the expected loss given default (ELGD) associated with the zero coupon debt contract as

$$ELGD = E^P[\text{Max}(0, B - V_T)I(A_T < D)], \quad (7)$$

and the default probability of the borrower (PD), under probability P , as

$$PD = E^P[I(A_T < D)]$$

Given these definitions, we can write (6) as

$$E^P[F(A_T, V_T, T)] = B - ELGD \cdot PD. \quad (8)$$

Analogously to (1), we can value any payoff by taking its discounted expectation under a martingale probability measure Q . Evaluating (7) and (8) under a risk neutral measure yields the value of the risky zero coupon debt as (see Appendix for the proof).

Proposition 1. *The value of the risky zero coupon debt is*

$$F(A_0, V_0, 0) = B(r, T) \left\{ 1 - L(b, \sigma_V^2, \rho, T; Q) Q(d, \sigma_A^2, T) \right\}, \quad (9)$$

where

$$B(r, T) = B e^{-rT}$$

$$L(b, \sigma_V^2, \rho, T; Q) =$$

$$\frac{\int_{-\infty}^{N^{-1}[Q]} \left\{ N \left[\frac{h_2 - \rho y}{\sqrt{1 - \rho^2}} \right] - \frac{1}{b} e^{-\frac{1}{2} \rho^2 \sigma_V^2 T + \rho \sigma_V \sqrt{T} y} N \left[\frac{h_2 - (1 - \rho^2) \sigma_V \sqrt{T} - \rho y}{\sqrt{1 - \rho^2}} \right] \right\} n(y) dy}{Q}$$

$$Q(d, \sigma_A^2, T) = N[h_1(d, \sigma_A^2, T)]$$

$$b = \frac{B}{V_0 e^{rT}}$$

$$d = \frac{D}{A_0 e^{rT}}$$

$$h_1(d, \sigma_A^2, T) = \frac{\ln(d) + 1/2 \sigma_A^2 T}{\sigma_A \sqrt{T}}$$

$$h_2(b, \sigma_V^2, T) = \frac{\ln(b) + 1/2 \sigma_V^2 T}{\sigma_V \sqrt{T}}$$

where $N(\cdot)$ is the standard normal distribution function, and $n(\cdot)$ is the standard normal density function.

The solution (9) is not quite an analytic formula in that it involves a one-dimensional integral which has to be evaluated numerically. Such one-dimensional quadrature is not computationally intensive, though, and can be performed very quickly using some of the well-known numerical integration schemes (see e.g. Press *et al.*, 1992).

Formula (9) is exactly of the same form as (1). In the formula (9), $B(r, T)$ is the present value of B to be received at time T , or the price of a risk free discount bond. This is multiplied by the expected payoff, per unit face value, of the debt at maturity. This expected payoff is

just the promised unit payment, from which the product of the ELGD and the default probability is deducted, both calculated under the martingale measure. The function $L(b, \sigma_V^2, \rho, T; Q)$ is the ELGD, expressed as a percentage of the face value of the debt. This is a function of the quasi loan-to-value ratio, b , as well as the parameters of the collateral value process. The function $Q(d, \sigma_A^2, T)$ is the default probability of the counterparty under the martingale measure. This is a function of the quasi leverage ratio of the firm, d , as well as the volatility of the asset value process.

The roles of the stochastic variables A and V are separated in a useful manner in our model. First, the default probability is only a function of the parameters of the asset value process (including d), and does not depend on the values of the collateral value parameters, nor on the particular debt contract's loan to value ratio b . Second, the ELGD is a function of the parameters of the asset value process only through Q , the quasi default probability. Thus it is possible to implement the model and to calculate the ELGD with external estimates of default probabilities, similarly to reduced form debt pricing models, without having to estimate the asset value parameters in the first place. The parameters of the collateral value process as well as the correlation between the collateral and the default determining variable could be estimated using appropriate proxy variables such as stock market industry indexes, or real estate indexes in the case of real estate collateral.

The debt value satisfies two useful homogeneity properties, which we record in the following. The proofs can be carried by substitution into the debt value formula (9).

Proposition 2. *The value of the risky zero coupon debt given in (9) is*

- (i) *homogenous of degree 1 in V_0 and B ,*
- (ii) *homogenous of degree 0 in A_0 and D .*

The fact that debt value is linearly homogenous in the face value of the debt and the value of the collateral is an implication of the linearity of pricing: as long as the loan-to-value ratio remains unchanged, the terms of the loan do not change in relative terms. That debt value is homogenous of degree zero in the values of the default determining variable and the default boundary implies that debt value only depends on the ratio of the two variables. This supports

our interpretation of A and D as a stylized description of a mechanism for creating a term structure for default probabilities, rather than as concrete asset values.

The debt value in (9) has an intuitive form, which becomes even simpler if we assume that the collateral value and the default determining variable are uncorrelated. Evaluating (9) with $\rho = 0$ yields the following expression for the debt value

$$F(A_0, V_0, 0) = B(r, T) - E^Q[I(A_T < D)]BS^{Put}(V_0, B, \sigma_V^2, T), \quad (10)$$

where BS^{Put} is the Black and Scholes value of a put option on V , with strike B . Equation (10) shows that the value of risky debt in this case is the value of the corresponding riskless debt, deducted by the default probability of the firm times the value of a put option on the collateral value. Since default probability is less than 1, the risky debt is more valuable than the portfolio consisting of the riskless debt and a short position in the put option on the collateral value. This holds even if the collateral value and the asset value are positively correlated. The holder of the risky debt is hurt by low collateral values only when the firm actually defaults. As far as the correlation between collateral and asset values is less than perfect, defaults do not always take place when collateral values are low.

Merton's original model of risky debt can be obtained as a special limit of our model, which can be achieved by merging the asset value and the collateral value processes appropriately. Our model thus incorporates the Merton model as a special case, and is therefore a genuine extension. We record this in the next proposition.

Proposition 3. *Merton's model of risky debt, where the value of a zero coupon bond is given by*

$$F^M(A_0, T) = B(r, T) \left\{ N[-h_1(d, \sigma_A^2, T)] + \frac{1}{d} N[h_1(d, \sigma_A^2, T) - \sigma_A \sqrt{T}] \right\},$$

is obtained as a limit of the debt value in (9) when $A_0 = V_0$, $B = D$, and the collateral and asset value parameters coincide, so that $\rho = 1$.

4 Sensitivity analysis

Formula (9) shows that the value of risky debt depends on the collateral value parameters only through their effect on the ELGD. Since the ELGD has a negative sign in the debt value formula, the partial derivatives of debt value with respect to the collateral value parameters are of opposite sign than the corresponding partial derivatives of the ELGD. In this section we therefore only analyse the sensitivities of the ELGD. The brute force way to derive these sensitivities would be to analyse the explicit solution for $L(\cdot)$ in (9). A perhaps easier way to derive the sensitivities is based on manipulating the definition of the ELGD, (7), as

$$\begin{aligned}
 ELGD &= E^P [Max(0, B - V_T) I(A_T < D)] \\
 &= E^P [Max(0, B - V_T) I(A_T < D)] / E^P [I(A_T < D)] \\
 &= E^P [Max(0, B - V_T)] + \frac{Cov^P [Max(0, B - V_T), I(A_T < D)]}{E^P [I(A_T < D)]}
 \end{aligned} \tag{11}$$

The last row in (11) shows that the ELGD can be expressed as the sum of the expected terminal payoff from a (long) put option on the collateral value, with strike equal to the face value of the debt, and a second term involving the covariance of the put option payoff with the indicator function of default. If the default determining variable and the collateral value are uncorrelated, then this term is zero, so that the ELGD equals the first term on the last row in (11). As the value of a put is an increasing function of the volatility of its underlying asset, the ELGD is the higher, the higher is the volatility of the collateral.

The functions $Max(0, B - V_T)$ and $I(A_T < D)$ are both non-increasing in V and A , respectively. This implies that when the correlation between A and V is positive, the covariance between these functions (viewed as random variables) is also positive. Therefore under positive correlation between V and A , collateral volatility contributes positively to the ELGD also via the second term in (11). Moreover, the correlation between $Max(0, B - V_T)$ and $I(A_T < D)$ increases with increasing correlation between V and A , so that the ELGD is also increasing with respect to the correlation parameter. These conclusions are in line with immediate intuition.

Numerical examples of the behaviour of the *ELGD*, expressed as a percentage of the face value of debt, are presented in figures 1 and 2. Figure 1 shows the ELGD as a function of

collateral value volatility and the correlation parameter, for a firm with 0.5% default probability over a one year horizon, and for a debt with one year maturity and initial loan-to-value ratio of 1. As proved above, the ELGD is increasing in the volatility and the correlation parameters. The variation of the ELGD as a function of these parameters appears economically quite significant. Given a correlation of 40%, doubling the volatility from 20% to 40% nearly doubles the ELGD from 22% to 39 %. We also note that in the presence of zero collateral value volatility, the ELGD in Figure 1 equals zero since the initial loan-to-value ratio equals 1.

As a reservation to these result, we note that we have not varied the collateral value drift as a function of its volatility and correlation. We simply assumed the drift to equal the riskless rate. For long maturity credits in particular, the effects of the drift on the ELGD can be substantial. Asset pricing theories suggest that the collateral value drift should be related to the collateral value volatility and the correlation. Stocks with high volatility would have high expected growth rates as well, *ceteris paribus*, which would produce a dampening effect on the conditional loss expectation. In principle, all the collateral value parameters – its drift, volatility and the correlation with asset value – should be jointly estimated by taking into account the asset pricing relationships which are not explicitly considered in our model.

Figure 2 illustrates the relationship between the ELGD and the probability of default. We observe that for positive correlation between collateral value and the firm asset value, the *ELGD* is the lower, the higher is the likelihood of default. This might at first appear counterintuitive. The logic behind this *ceteris paribus* result is that the lower is the probability of default, the more the firm asset value has to decline before default will occur. Positive correlation between the asset value and the collateral value in turn implies that, for a given initial collateral value, the future collateral value in the event of default will on average be the lower, the lower is the initial default probability. Hence the negative relationship between initial default probability and the ELGD. We also note from Figure 2 that this relationship vanishes for zero correlation between the collateral and the firm value. Although the effect does not appear economically very significant for a wide range of parameter values, in some cases it may not be without significance. Figure 2 e.g. shows that for a collateral volatility of 15% and correlation with the asset value of 60%, the ELGD estimate for a top quality credit

(one year default probability close to zero) is about twice that of a quite low quality credit (default probability higher than 10%) with similar collateral position.

5 Loan-to-value ratios

As a second application of our model, we analyse several rules that banks could use in determining limits on loan-to-value ratios (the ratio of the face value of the debt to the collateral value pledged against the debt, B/V_0 in our model) in lending. Limits on loan-to-value ratios can at least partially substitute for risk sensitive pricing of loans, in that sufficient protective collateral limits potential losses in the event of default, and thus makes loans ‘sufficiently riskless’ to be eligible for a uniform, risk insensitive pricing. Loan-to-value ratios, or implied collateral haircuts, set according to such objectives should itself be sensitive to both the default risk of the counterparty, and the risk characteristics of the collateral value.

In order to implement the previous idea, a definition of ‘sufficiently riskless’ is required. Clearly value could be used as a criterion: one might require that the value of the loan, calculated from (9), implies a credit spread of 10 basis points or less. On the basis of this uniform value criterion, one could calculate limits on loan-to-value ratios as a function of the default probability and the parameters of the collateral value, then apply those limits in lending, and price all loans uniformly at 10 basis points over the riskless rate⁴. We present loan-to-value ratios calculated according to this rule in Table 1A.

We note from Table 1A that the loan-to-value ratios calculated according to the uniform value criterion decline with increasing collateral value volatility, decline with increasing correlation between collateral and asset value, and decline with increasing default probability of the borrower. Loan-to-value ratios also usually decline as the maturity of the contract is increased (in which case the cumulative default probability is also increased appropriately). The opposite happens, however, when the volatility of the collateral value is very low. This is because in these cases the collateral value drift begins to dominate the collateral value volatility as maturity lengthens.

⁴ In practitioner parlance, this would be called ‘lending on collateral’.

The uniform value criterion to setting loan-to-value ratios essentially takes a loan value as given, and maps it into a corresponding loan-to-value ratio, or equivalently, a collateral requirement. Solving for the loan-to-value ratio requires valuing the loan in the process. Sometimes this can be a laborious task that one would like to avoid, in particular since valuation may require the estimation of non-observable quantities such as risk-neutral probabilities of default. We are interested in finding out whether there could be other rules for setting loan-to-value ratios which would depend on the statistical model of the collateral and the asset value only, and which would produce results which are qualitatively similar to the ones based on the uniform value criterion? We test two probabilistic criteria below.

A Value-at-Risk type rule based on the conditional probability of loss is the following: allow loan-to-value ratios up to the point where the conditional probability of the collateral value falling below the face value of the debt in the event of default is at most X%. Table 1B presents loan-to-value ratios based on this rule for different values of the collateral value parameters. We observe that under this rule, loan-to-value ratio declines as either the collateral value volatility or the correlation between collateral value and asset value increases. Also, loan-to-value ratios decline as the maturity of the loan increases, unless the volatility of the collateral value is very low. This behaviour is in line with the sensitivity results obtained from the uniform value rule. Contrary to the uniform value rule, however, the loan-to-value ratio increases as the credit quality of the firm declines, given that the correlation between collateral and asset value is positive. In some cases this effect can be quite significant. This result is analogous to the negative relation between default probability and the ELGD observed previously. Yet the result is counterfactual, since casual empirical evidence suggests that loan-to-value ratios actually decline with increases in the default probability of the borrower. More seriously, this kind of limit system violates basic economic principles of value, in that if pricing is not risk sensitive, one should clearly not lend more to a lower quality customer, provided he can pledge the same collateral as a better quality customer. This odd behaviour is of course due to the fact that pricing cannot be based on information which is conditioned on default having taken place alone. The probability of default must clearly be reflected in the loan-to-value criterion, if this is to be substituted for risk sensitive pricing.

We have tested an alternative criterion based on unconditional probability of loss, stating: allow loan-to-value ratios up to the point where the unconditional probability of a credit loss is at most X%. Interpreted differently, the probability that the customer defaults and that, simultaneously, the collateral value falls below the nominal loan amount must not exceed a certain percentage. A immediate implication of this criterion is that any loan-to-value ratio is acceptable with borrowers whose default probability is below the unconditional loss probability which is used in the criterion⁵. In table 1C, we illustrate loan-to-value ratios determined based on this criterion. The criterion, with 0.1% as the maximal allowed probability of loss, produces both infinite loan-to-value ratios, as well as loan-to-value ratios above one hundred percent. In contrast to the previous criterion, however, the loan-to-value ratios calculated based on this criterion decline as the default probability of the borrower increases. This criterion in fact displays all the same comparative statics as the uniform price rule, and is therefore the more recommendable of the two probabilistic criteria analysed here.

6 Conclusions and discussion on the applicability under Basel II

In essence this paper is an attempt to provide insights to the estimation of a central credit risk quantity, loss-given-default, that is often treated as a constant for practical reasons, but that really is stochastic because of the uncertainty regarding the future value of collateral. Although our model is rather simple and stylized, we believe it can be a useful quantitative tool. Be it the estimation of the expected loss-given-default to be used in credit risk models, or determination of loan-to-value ratios for lending guidelines, the model helps to understand which parameters are relevant to consider and what are their likely effects. Moreover, we believe that our framework could be useful in estimating losses-given-default in the advanced internal ratings based approach of the new Basel capital accord.

Although the document of Basel Committee (2001) already contains many highly structured proposals for determining banks' minimum capital requirements for credit risk, some of them leave considerable room for banks themselves to decide what particular techniques to employ. The prime proposal in this category is the advanced internal ratings

⁵ In bank practitioners' parlance, this is sometimes called as 'lending against name'.

based approach for corporate credit exposures, in which banks would be allowed to provide their own loss-given-default estimates subject to certain qualitative minimum requirements. In forming these estimates, any type of collateral could in principle be considered. The loss concept would be that of expected loss-given-default (Basel Committee, 2001, paragraph 326). We suggest that our model could be applied for this purpose. As we have shown above, our method can be used to answer the question of how much is one expected to lose, in the event of default, from a given exposure which is protected by a given amount of collateral. Our approach takes this estimation problem down to the fundamental structural parameters, namely the drift and volatility of the collateral value, its correlation with the obligor's default probability, and the obligor's default probability. Of course a prerequisite for the acceptance of any internal model is the validation of the model-based estimates against empirical data.

The advanced IRB approach is not the only potential application of our framework within the new Basel Accord. In the so called foundation IRB approach, Basel Committee (2001) has suggested collateral haircuts for various asset types. It would be interesting to study whether these are consistent with those that can be obtained from our model. Moreover, the January 2001 proposal of the Basel Committee (2001) was rather short on the treatment of retail portfolios. Our framework might be useful in assessing losses-given-default in retail loans often backed by various types of physical collateral as well.

The Basel Committee's proposal has provoked critique (e.g. Frye, 2001) for ignoring the systematic risk component of loss-given-default, which results from the exposure of collateral values to the common risk factors driving defaults. We believe that this criticism could be at least partly accounted for within the current Basel proposal by using appropriate collateral haircuts which take into account the systematic risk. A collateral haircut percentage y is by definition a number which satisfies

$$ERGD = (1 - y)V_0$$

where ERGD stands for the expected recovery given default. Within our model, the ERGD and the ELGD of a zero coupon bond are related by $ERGD + ELGD = B$, so that our ELGD estimates can easily be mapped into haircut percentages. When e.g. the face value of the debt equals the current collateral value ($B = V_0$), the ELGD percentages in figures 1 and 2 directly correspond to the haircut percentage.

A more subtle point concerns the correct definition of the event on which the expectation should be conditioned in the calculation of the ELGD. We have calculated ELGD conditional on the borrower's own default. In a portfolio context, where the task is to allocate capital which is estimated as a given percentile point of a portfolio distribution, to individual obligors and exposures, the correct conditioning event is one where the portfolio value is equal to the given percentile point in its distribution. This event is much more complicated than a default of a single counterparty, since in general there are several possible combinations of defaults by different obligors which all give rise to the same portfolio loss. Praschnik et al. (2001) discuss one way to numerically calculate such expectations.

Appendix

The value of risky debt is obtained by taking a discounted expectation, under a martingale measure Q , of the payoff at T which is given by equation (5) in the text,

$$\begin{aligned} F(A_0, V_0, 0) &= e^{-rT} E^Q [F(A_T, V_T, T)] \\ &= B(r, T) \left\{ 1 - \frac{1}{B} E^Q [Max(0, B - V_T) I(A_T < D)] \right\}, \end{aligned} \quad (A1)$$

where $B(r, T) = Be^{-rT}$ is the present value of the riskless payoff B to be received at time T in the future. In the presence of a stochastic riskless term structure, the function $B(r, T)$ would be the price of a riskless discount bond under the chosen term structure model. The challenge is to evaluate the term $E^Q [Max(0, B - V_T) I(A < D)]$. For this purpose, let us denote the joint density function of A_T and V_T as $f(A, V)$, the density function of V_T , conditional on A_T , as $f_{V|A}(A, V)$, and the marginal density function of A_T as $f_A(A)$. The expectation can then be evaluated as follows (we have dropped time-subscripts)

$$\begin{aligned} &\frac{1}{B} E^Q [Max(0, B - V) I(A < D)] \\ &= \frac{1}{B} \int_0^\infty \left(\int_0^\infty Max(0, B - V) I(A < D) f(A, V) dV \right) dA \\ &= \frac{1}{B} \int_0^D \left(\int_0^B (B - V) f_{V|A}(A, V) dV \right) f_A(A) dA \\ &= \int_0^D \left(\int_0^B f_{V|A}(A, V) dV \right) f_A(A) dA - \frac{1}{B} \int_0^D \left(\int_0^B V f_{V|A}(A, V) dV \right) f_A(A) dA \\ &= \int_{-\infty}^{h_1} \left(\int_{-\infty}^{h_2} f_{z|y}(y, z) dz \right) f_y(y) dy \\ &\quad - \frac{1}{b} \int_{-\infty}^{h_1} \left(\int_{-\infty}^{h_2} \exp\left(-\frac{1}{2}\sigma_V^2 T + \sigma_V \sqrt{T} z\right) f_{z|y}(y, z) dz \right) f_y(y) dy \end{aligned} \quad (A2)$$

where y and z are jointly standard normally distributed random variables with correlation ρ ,

$b = \frac{B}{V_0 e^{rT}}$ is the quasi loan-to-value ratio of the zero coupon debt, $d = \frac{D}{A_0 e^{rT}}$ is the quasi

leverage ratio of the firm, and

$$h_1(d, \sigma_A^2, T) = \frac{\ln(d) + 1/2 \sigma_A^2 T}{\sigma_A \sqrt{T}},$$

$$h_2(b, \sigma_V^2, T) = \frac{\ln(b) + 1/2 \sigma_V^2 T}{\sigma_V \sqrt{T}}.$$

From the properties of the bivariate normal distribution, we know that conditional on y , z is normally distributed as $z|y \sim N(\rho y, 1 - \rho^2)$. Using this knowledge, the first term in the final expression in (A2) can be written as

$$\int_{-\infty}^{h_1} N\left(\frac{h_2 - \rho y}{\sqrt{1 - \rho^2}}\right) n(y) dy. \quad (\text{A3})$$

The second term in the final expression in (A2) can be completed into a square prior to integrating. The expression inside the inner integral becomes

$$\begin{aligned} & \exp\left(-\frac{1}{2} \sigma_V^2 T + \sigma_V \sqrt{T} z\right) f_{z|y}(y, z) \\ &= \exp\left(-\frac{1}{2} \rho^2 \sigma_V^2 T + \rho \sigma_V \sqrt{T} y\right) \frac{1}{\sqrt{2\pi(1 - \rho^2)}} \exp\left(-\frac{1}{2(1 - \rho^2)} \left(z - (\rho y + (1 - \rho^2) \sigma_V \sqrt{T})\right)^2\right) \end{aligned}$$

Substituting this expression into the last row in (A2) yields

$$\frac{1}{b} \int_{-\infty}^{h_1} \exp\left(-\frac{1}{2} \rho^2 \sigma_V^2 T + \rho \sigma_V \sqrt{T} y\right) N\left(\frac{h_2 - (1 - \rho^2) \sigma_V \sqrt{T} - \rho y}{\sqrt{1 - \rho^2}}\right) n(y) dy. \quad (\text{A4})$$

Combining (A2), (A3) and (A4) gives

$$\begin{aligned} & \frac{1}{B} E^Q[\text{Max}(0, B - V) I(A < D)] = \int_{-\infty}^{h_1} N\left(\frac{h_2 - \rho y}{\sqrt{1 - \rho^2}}\right) n(y) dy \\ & - \frac{1}{b} \int_{-\infty}^{h_1} \exp\left(-\frac{1}{2} \rho^2 \sigma_V^2 T + \rho \sigma_V \sqrt{T} y\right) N\left(\frac{h_2 - (1 - \rho^2) \sigma_V \sqrt{T} - \rho y}{\sqrt{1 - \rho^2}}\right) n(y) dy \end{aligned} \quad (\text{A5})$$

Noting that $Q(d, \sigma_A^2, T) = N[h_1(d, \sigma_A^2, T)]$, substituting (A5) into (A1), and using the definition of conditional expectation, we get equation (9) in the text.

References

- Anderson, R., S. Sundaresan, 1997, Design and valuation of debt contracts, *Review of Financial Studies*, 9, 37-68.
- Basel Committee for Banking Supervision, January 2001, The New Basel Capital Accord, Bank for International Settlements, Basel.
- Black, F., J.C. Cox, 1976, Valuing corporate securities: Some effects of bond indenture provisions, *The Journal of Finance*, 31, 351-367.
- Black, F., M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy*, 81, 637-54.
- Briys, E., F. de Varenne, 1997, Valuing risky fixed rate debt: an extension, *Journal of Financial and Quantitative Analysis*, 32, 239-248.
- Credit Suisse Financial Products, 1997, *CreditRisk⁺ - A Credit Risk Management Framework*, 1997.
- Duffie, D., K.J. Singleton, 1999, Modelling term structures of defaultable bonds, *Review of Financial Studies*, 12, 4, 1999.
- Fan, H., S. Sundaresan, 2000, Debt valuation, renegotiation, and optimal dividend policy, *Review of Financial Studies*, 13, 1057-1099.
- Frye, J., 2000a, Collateral damage, *Risk*, April, 91-94.
- Frye, J., 2000b, Depressing recoveries, *Risk*, November, 106-111.
- Frye, J., 2001, Weighting for risk, *Risk*, May, 91-94.
- Jarrow, R.A., S.M. Turnbull, 1995, Pricing derivatives on financial securities subject to credit risk, *The Journal of Finance*, 50, 53-86.
- Jarrow, R.A., D. Lando, S.M. Turnbull, 1997, A Markov model for the term structure of credit risk spreads, *Review of Financial Studies*, 10, 2, 481-523.
- J.P. Morgan, 1997, *CreditMetricsTM - Technical Document*, New York.
- Kealhofer, S., 1993, Portfolio Management of Default Risk, KMV Corporation, San Francisco.

- Kim, I., K. Ramaswamy, S. Sundaresan, 1993, Does default risk in coupons affect the valuation of corporate bonds: A contingent claims model, *Financial Management*, 22, 117-131.
- Leland, H.E., 1994, Risky debt, bond covenants, and optimal capital structure, *The Journal of Finance*, 49, 1213-1252.
- Leland, H.E., K. Toft, 1996, Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads, *The Journal of Finance*, 51, 987-1019.
- Longstaff, F.A., E.S. Schwartz, 1995, A simple approach to valuing risky fixed and floating rate debt, *The Journal of Finance*, 50, 789 – 819.
- Mella-Barral, P., W. Perraudin, 1997, Strategic debt service, *The Journal of Finance*, 52, 531-556.
- Merton, R.C., 1974, On the pricing of corporate debt: the risk structure of interest rates, *The Journal of Finance*, 29, 449-470.
- Praschnik, J., G. Hayt, A. Principato, 2001, Calculating the contribution, *Risk*, October, S25-S27.
- Press, W.H., S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, 1992, *Numerical Recipes in C*. 2. ed., Cambridge University Press.
- Schleifer, A. and R. Vishny, 1992, Liquidation Values and Debt Capacity: A Market Equilibrium Approach, *The Journal of Finance*, 47, 1343-1366.

Figure 1. The ELGD as a function of collateral volatility and correlation

Constant parameters: (one year) default probability 0.5%, loan-to-value ratio 1, debt maturity 1 year, riskless rate 5%, collateral value drift 5%, residual recovery rate 0%.

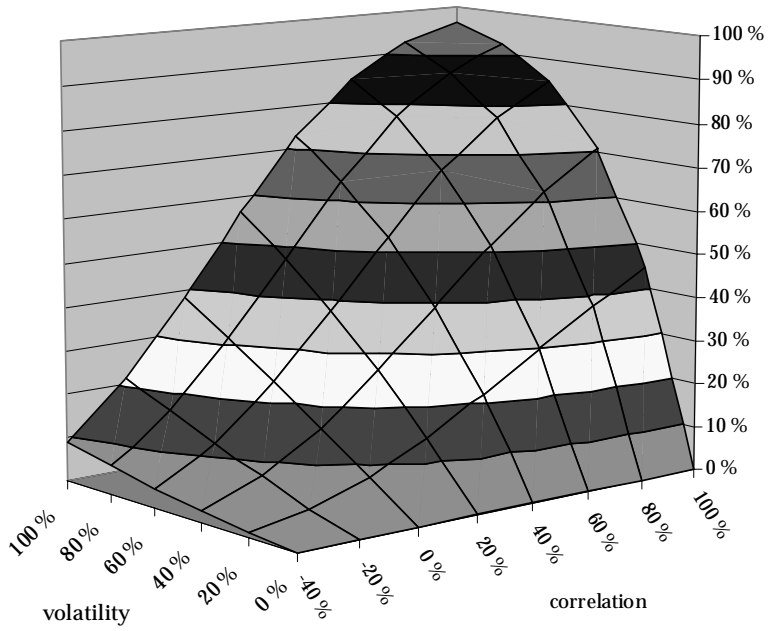


Figure 2. The ELGD as a function of default probability

Constant parameters: loan-to-value ratio 1, debt maturity 1 year, riskless rate 5%, collateral value drift 5%, collateral value volatility 15%, residual recovery rate 0%.

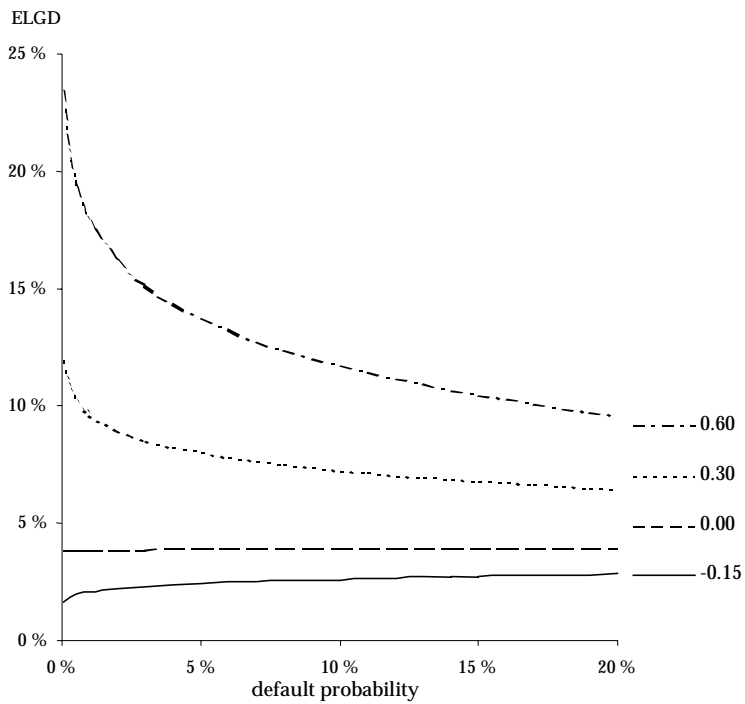


Table 1. Loan-to-value ratios

The (cumulative) one/three-year default probabilities for A, BB, and B rated counterparties are 0.03%/0.22%, 1.32%/6.01%, and 5.58%/15.6%, respectively. These are obtained by assuming a time-homogenous transition matrix of annual rating transition probabilities, based on Standard and Poor's default statistics (see J.P. Morgan, 1997, table 6.11). Risk free rate is 5% and collateral value drift is 5%.

1A: Loan-to-value ratios based on uniform value criterion

Highest ratio of loan amount to current collateral value such that the yield spread on the loan, over the riskfree rate, is less than 0.1%.

correlation	volatility rating	5 %		15 %		30 %	
		1-yr	3-yr	1-yr	3-yr	1-yr	3-yr
0 %	A	*	*	*	*	*	*
	BB	113	120	109	99	94	67
	B	105	111	93	87	72	52
30 %	A	*	*	*	*	*	*
	BB	109	113	97	87	75	51
	B	101	107	85	78	61	42
60 %	A	*	*	*	*	*	*
	BB	105	109	87	77	62	41
	B	99	105	80	72	54	36

* All loan-to-value ratios satisfy the criterion.

1B: Loan-to-value ratios based on conditional loss probability

Highest ratio of loan amount to current collateral value such that the conditional probability of loss, given bankruptcy, is less than 5.0%.

correlation	volatility rating	5 %		15 %		30 %	
		1-yr	3-yr	1-yr	3-yr	1-yr	3-yr
0 %	A	97	100	81	73	61	43
	BB	97	100	81	73	61	43
	B	97	100	81	73	61	43
30 %	A	92	93	69	59	45	27
	BB	93	96	73	64	50	33
	B	94	97	75	66	53	36
60 %	A	87	88	61	49	35	19
	BB	90	92	67	57	42	27
	B	92	94	71	62	46	31

* All loan-to-value ratios satisfy the criterion.

1C: Loan-to-value ratios based on unconditional loss probability

Highest ratio of loan face value to current collateral value such that the unconditional probability of loss is less than 0.1%.

correlation	volatility rating	5 %		15 %		30 %	
		1-yr	3-yr	1-yr	3-yr	1-yr	3-yr
0 %	A	*	114	*	109	*	98
	BB	95	95	84	63	66	34
	B	94	93	74	58	53	27
30 %	A	*	106	*	86	*	62
	BB	94	91	77	56	53	26
	B	90	90	70	53	46	22
60 %	A	*	98	*	67	*	36
	BB	90	89	71	52	45	22
	B	89	89	67	51	41	21

* All loan-to-value ratios satisfy the criterion.

Essay 5: On the markup interpretation of optimal stopping rules^{*}

Abstract

In the context of the standard irreversible investment problem, Dixit et al. (1999) derived a first-order condition for optimal stopping rules that has a markup interpretation familiar from classical producer theory. We show here that optimal policies can be characterized through first-order conditions that have the markup interpretation also in stopping models with flow payoffs and multiple (an infinite sequence of) stopping decisions. We do this by analyzing two well known models in the real options literature: a model of abandonment where the objective is the present value of a flow payoff, and Dixit's (1989) regenerative model of investment and abandonment. The markup formulas follow from conveniently representing the objective function as a function of the stopping barriers, taking advantage of renewal arguments. In many cases this approach is equally tractable as the dynamic programming approach, provides a different economic intuition into the solution, and is pedagogically very appealing.

Keywords: optimal stopping, investment, abandonment, markup

JEL classification: D21, G31

1 Introduction

The standard approach to optimal stopping is via dynamic programming/variational inequalities, as illustrated e.g. in Bensoussan and Lions (1982) or Oksendal (1995). A textbook example of an optimal stopping problem whose dynamic programming differential equation can be solved explicitly is the plain vanilla irreversible investment problem first solved by McDonald and Siegel (1986) and extensively analyzed by Dixit and Pindyck (1994). It is well known that this particular problem can also be solved using a direct approach, whereby the objective function is represented in terms of the unknown stopping

^{*} I thank Luis Alvarez for very helpful comments that led to numerous improvements in this paper.

barrier, and then optimized with respect to the location of the barrier. Moreover, as noted by Dixit et al. (1999), the direct approach yields a first-order condition for the optimal investment barrier that has exactly the same form as the optimal pricing rule of a firm that faces a downward sloping demand curve, generally referred to as the markup pricing rule. Dixit et al. hence refer to this first-order condition as the markup interpretation of optimal investment rules.

The markup first-order condition obtained by Dixit et al. (1999) follows automatically from the representation of the objective in terms of the unknown barrier. The direct approach therefore provides a different economic intuition into the solution relative to dynamic programming, an intuition that most economists may be more familiar with and which could be lost if the problem were attacked through solution of the HJB equation. The direct approach also has other merits relative to dynamic programming. First, solution of the optimal policies in the direct approach reduces to solving a system of algebraic equations, as opposed to solving a differential equation. Second, the system of algebraic equations in the direct approach does not involve other unknowns besides the stopping barrier(s). The dynamic programming approach eventually leads to first-order conditions for the barrier(s) as well, but this system is of higher dimension since it contains unknown coefficients of the general solution to the HJB equation.

In this paper, we apply the direct approach to some straightforward variations and generalizations of the basic irreversible investment problem, and obtain first-order conditions that have ‘classical’ mark-up interpretations in terms of relevant elasticities. We first study a problem of optimal abandonment of a firm generating a flow of rewards. We find this particular model well suited for demonstrating the simplicity of the technique, for regenerating the closed form solutions obtained earlier based on dynamic programming, and for showing that with the direct approach, variations in the problem structure can be handled very efficiently. We then apply the direct approach to a regenerative stopping problem with both flow rewards and lump-sum payments at stopping times, i.e. to Dixit’s (1989) model of sequential investment and abandonment. As for both of these models, we claim that the direct approach is conceptually simpler and (at least) equally tractable as dynamic programming, and yields first-order conditions with clear economic interpretations.

A necessary condition for the direct approach to work is that the objective function be represented as a function of the stopping barriers in terms of components that are suitable for explicit computation. As for the standard irreversible investment problem, the representation obtained by Dixit et al. (1999) is a product of two terms, both of which are functions of the barrier. The essential argument that Dixit et al. (1999) need to obtain this representation is the time-homogeneity of the problem. The generalizations that we analyze preserve time-homogeneity, and we assume the state variable to be a strong Markov process. We use renewal arguments which take advantage of just these properties to obtain convenient 'product' representations of the objective. We acknowledge that other methods could be applied to represent the objective as a function of the barriers. A representation based on Green functions has been used by Alvarez (1998, 1999, 2001) to solve one-off stopping problems using the direct approach. The first-order conditions obtained by Alvarez however do not have the mark-up character. In regenerative problems, in particular, it appears that renewal techniques are essential for representing the objective in terms of components that are amenable to explicit computation (these component solutions may be obtained using the method of Green functions, but as for the models analyzed in this paper, the technique is unnecessarily general). Related techniques have also been utilized in the (s,S) inventory literature (e.g. Porteus, 1971, 1972, Harrison, 1985, Bentolila and Bertola, 1990), but these papers have not been concerned with the mark-up interpretation of the first-order conditions. The paper most related to ours is Sodal (2002), which analyzes the stochastic rotation problem of Faustmann using the same methods that we use in this paper. The model in Section 5 of this paper is a generalization of the model analyzed by Sodal.

The paper is organized as follows. Section 2 presents a brief summary of Dixit et al.'s (1999) results. Section 3 introduces an optimal abandonment problem with flow rewards and derives the markup first-order condition. Section 4 presents explicit solutions to this problem under two different stochastic processes for the state variable, and demonstrates the efficiency of the direct method in accommodating a variation in the problem structure. Section 5 applies the direct approach to a model of joint investment and abandonment, derives first-order conditions analogous to those in the simpler models, and obtains the objective function in closed form in the case of geometric Brownian motion price dynamics. Section 6 concludes.

2 An overview of Dixit et al. (1999)

Dixit et al. (1999) developed the markup interpretation in the context of the standard irreversible investment problem first analyzed by McDonald and Siegel (1986). The state variable P representing the net present value of an investment project is assumed to follow a time-homogenous diffusion process (not necessarily geometric Brownian motion). The objective is to maximize the value V of the option to undertake this investment against a fixed investment cost I , by optimally choosing the exercise time of the investment option

$$V(P) = \sup_{\tau} E_P \left[e^{-\rho\tau} (P_{\tau} - I) \right] \quad (1)$$

where the optimization is over all stopping times τ of the filtration generated by the state variable P . The discount rate ρ is taken to be a constant. Since the optimal stopping times in problem (1) are hitting times of a fixed barrier b , the problem reduces to

$$V(P) = \sup_b E_P \left[e^{-\rho\tau_b} (b - I) \right] = \sup_b E_P \left[e^{-\rho\tau_b} \right] (b - I) = \sup_b L(P, b)(b - I) \quad (2)$$

where τ_b denotes the first hitting time of P to the barrier b from below, defined as $\tau_b = \inf\{t \geq 0 : P_t \geq b\}$, and $L(P, b)$ denotes the Laplace transform of τ_b , given that $P_0 = P$. The elasticity of the Laplace transform of the hitting time, with respect to the barrier b , is defined by

$$\varepsilon_{L_b}(b) = -\frac{L_b(P, b)b}{L(P, b)}. \quad (3)$$

Dixit et al. show that this elasticity is independent of the starting value P , although in general the elasticity depends on the value of the barrier b . The first-order condition to (2) then takes a specially intuitive form when expressed in terms of the elasticity,

$$\frac{b^* - I}{b^*} = \frac{1}{\varepsilon_{L_b}(b^*)}, \quad (4)$$

which is identical in form to the markup pricing rule of a firm facing a declining demand curve. Indeed, interpret I as the marginal cost of the firm's output (assumed constant), b as the output price, and $L(P, b)$ as the demand curve facing the firm (which is declining in b , and increasing in P which may be interpreted as a shift parameter). Then equation (2) is a standard

one-period profit maximization problem, and (4) is the optimal pricing rule, ε_{L_b} in this case being the price elasticity of demand.

Returning to the optimal investment problem, rearranging (4) in terms of b^* yields an implicit expression for the optimal investment threshold in terms of the elasticity

$$b^* = \left(\frac{\varepsilon_{L_b}(b^*)}{\varepsilon_{L_b}(b^*) - 1} \right) I. \quad (5)$$

This is in general only an implicit expression since ε_{L_b} may depend on b . It is obvious from this expression that when the elasticity is positive and larger than 1 in absolute value, the optimal barrier b^* will exceed the break even value I .

For some typical diffusion processes used to model project values, explicit expressions for $L(P, b)$ are known. When P satisfies a geometric Brownian motion with parameters $\mu > 0$ and $\sigma > 0$, we have

$$L(P, b) = \left(\frac{P}{b} \right)^{\beta_1}, \quad (6)$$

where β_1 is the positive solution to the fundamental quadratic associated with geometric Brownian motion,

$$\frac{1}{2} \sigma^2 \beta(\beta - 1) + \mu\beta - \rho = 0,$$

and is larger than 1 in well behaved cases, i.e. when $\mu < \rho$. Substituting (6) into (3), we observe that the elasticity of the Laplace transform with respect to b is constant and equal to β_1 . Substituting this into (5), we obtain the familiar result for the optimal investment barrier

$$b^* = \left(\frac{\beta_1}{\beta_1 - 1} \right) I.$$

3 An abandonment problem

Our analysis here is in the context of a stopping problem which has earlier been studied based on dynamic programming methods by Dixit (1989) and Leland and Toft (1996), among

others. A firm with unit output and a constant unit cost C experiences a stochastic output price P , which follows a time-homogenous diffusion process which satisfies the strong Markov property. Hence the firm's instantaneous profit flow is $P_t - C$. The firm value is the present value of this profit flow, based on a constant discount rate ρ , allowing for the firm's option to irreversibly terminate its operations, should the profit flow become sufficiently negative. Again, we know from the problem structure that the optimal stopping times must coincide with first hitting times of a constant barrier. Denoting the stopping barrier by b , we write firm value as

$$V(P; b) = E_p \left[\int_0^{\tau_b} e^{-\rho t} (P_t - C) dt \right], \quad (7)$$

where τ_b denotes the first hitting time of P to the barrier b from above, defined by $\tau_b = \inf\{t \geq 0 : P_t \leq b\}$. We want to write this objective in a form that is suitable for explicit calculation, and then optimize the barrier by differentiating the resulting expression with respect to b . We first write (7) as

$$\begin{aligned} V(P; b) &= E_p \left[\int_0^{\tau_b} e^{-\rho t} P_t dt \right] - E_p \left[\int_0^{\tau_b} e^{-\rho t} C dt \right] \\ &= E_p \left[\int_0^{\infty} e^{-\rho t} P_t dt \right] - E_p \left[e^{-\rho \tau_b} \right] E_b \left[\int_0^{\infty} e^{-\rho t} P_t dt \right] \\ &\quad - E \left[\int_0^{\infty} e^{-\rho t} C dt \right] + E_p \left[e^{-\rho \tau_b} \right] E \left[\int_0^{\infty} e^{-\rho t} C dt \right] \end{aligned} \quad (8)$$

where we have used the fact P is assumed to be a strong Markov process. In order to economize the presentation, we define the following functions

$$W(P) = E_p \left[\int_0^{\infty} e^{-\rho t} P_t dt \right] \quad (9a)$$

$$W_0(C) = E \left[\int_0^{\infty} e^{-\rho t} C dt \right] \quad (9b)$$

$$L(P, b) = E_p \left[e^{-\rho \tau_b} \right]. \quad (9c)$$

Here as before, $L(P,b)$ is the Laplace transform of the first hitting time of P to the barrier b . Note, however, that although we use the same notation $L(P,b)$ for the Laplace transform in the optimal investment problem and in the optimal abandonment problem, the hitting times are not the same and therefore the Laplace transforms in the two models are two different functions.

Substituting the definitions (9a-c) into (8), we have

$$V(P;b) = (W(P) - W_0(C)) + L(P,b)(W_0(C) - W(b)). \quad (10)$$

This is a decomposition of the firm value into two terms. The component inside the first parentheses is the present value of the profit flow, given the current price, assuming that the firm is operated indefinitely. This term may be negative for sufficiently low P , and does not depend on b . The second component $L(P,b)(W_0(C) - W(b))$ is the value of the option to terminate operations irreversibly. The present value factor $L(P,b)$ is always non-negative, so that for the option value to be positive, $W(b)$ must be less than $W_0(C)$. When P is a growing process, this already implies that the optimal b must be less than C . Optimal choice of b maximizes the option value and hence the firm value. The first-order condition for optimality of b in (10) is

$$V_b(P;b^*) = L_b(P,b^*)(W(b^*) - W_0(C)) + L(P,b^*)W'(b^*) = 0. \quad (11)$$

We define the elasticity of the Laplace transform $L(P,b)$ with respect to b as in (3). Again, the elasticity only depends on b since similar reasoning as in Dixit et al. (1999) can be used to show that the elasticity must be independent on P . The first-order condition (11) now can be expressed as

$$\frac{W(b^*) - W_0(C)}{W'(b^*)b^*} = -\frac{1}{\varepsilon_{L_b}(b^*)}, \quad (12)$$

which is closely analogous to the markup formula (4) of the optimal investment problem. Moreover, when $W(b)$ is first-degree homogenous in b , as will be the case when P follows a geometric Brownian motion, we have $W(b) = W'(b)b$, so that equation (12) becomes

$$\frac{W(b^*) - W_0(C)}{W(b^*)} = -\frac{1}{\varepsilon_{L_b}(b^*)}.$$

This is like the markup formula (4) of the optimal investment problem, with the exception that the optimal stopping barrier for the value of the investment project b^* , and the investment cost I , have been replaced by the expected present values $W(b^*)$ and $W_0(C)$ of the price flow and the cost flow, as defined in (9a) and (9b). These expectations are conditional on the current value of the price level being at the optimal stopping barrier. Solving the previous equation in terms of $W(b^*)$ gives

$$W(b^*) = \left(\frac{\varepsilon_{L_b}(b^*)}{\varepsilon_{L_b}(b^*) - 1} \right) W_0(C), \quad (13)$$

which is completely analogous to (5).

There is one more form for the first-order condition that is very useful, in particular to show the analogy between the current problem where a single stopping decision is analyzed, and the more general problem in Section 5 where successive investments and abandonments are studied. Let us denote the present value of the undiscontinued profit flow as $\tilde{W}(P) = W(P) - W_0(C)$, and define its elasticity with respect to P as

$$\varepsilon_{\tilde{W}_P}(P) = \frac{\tilde{W}_P(P)P}{\tilde{W}(P)}.$$

Then we can write condition (12) as

$$\varepsilon_{\tilde{W}_P}(b^*) = \varepsilon_{L_b}(b^*). \quad (14)$$

This states that the elasticity of the present value of the profit flow equals the elasticity of the discount factor, when both are evaluated at the optimal exit barrier.

4 Explicit solutions

Applying formula (12) is very straightforward if explicit expressions for the functions $W(P)$, $W_0(C)$, and $L(P, b)$ defined in (9a-c) can be found. We show this under two assumptions concerning the dynamics of P which yield well known closed form solutions to (9a-c).

4.1 Geometric Brownian motion

Let P satisfy the stochastic differential equation

$$dP = \mu P dt + \sigma P dB ,$$

where B is a standard Brownian motion and μ and σ are positive parameters such that $\mu < \rho$. Then explicit solutions to (9a-c) are (see e.g. Dixit, 1993)

$$W(P) = \frac{P}{\rho - \mu} , \quad (15a)$$

$$W_0(C) = \frac{C}{\rho} , \quad (15b)$$

$$L(P, b) = \left(\frac{P}{b} \right)^{\beta_2} . \quad (15c)$$

Here β_2 is the negative solution to the fundamental quadratic associated with geometric Brownian motion,

$$\frac{1}{2} \sigma^2 \beta(\beta - 1) + \mu\beta - \rho = 0 .$$

Substituting (15c) into formula (3) for the elasticity, we get $\varepsilon_{L_b} = \beta_2 < 0$. Under geometric Brownian motion, the elasticity is constant as in the model of optimal investment. Moreover, since (15a) is first degree homogenous in P , we can utilize the markup formula (13) to determine the optimal abandonment barrier. Substituting in (15a), (15b), and β_2 for the elasticity, yields the optimal barrier

$$b^* = \left(\frac{\beta_2}{\beta_2 - 1} \right) \left(\frac{\rho - \mu}{\rho} \right) C .$$

Since $\beta_2 < 0$ and $\rho > \mu$, we have $b^* < C$. In the case of abandonment, the optimal barrier is below the cost rate, which is due to the option value of waiting. This is the same formula that could have been obtained based on dynamic programming methods.

4.2 Arithmetic Brownian motion

Now assume that P satisfies

$$dP = \mu dt + \sigma dB ,$$

where B is a standard Brownian motion and μ and σ are positive parameters. Then for the formulas (9a-c) we get

$$W(P) = \frac{P}{\rho} + \frac{\mu}{\rho^2}, \quad (16a)$$

$$W_0(C) = \frac{C}{\rho}, \quad (16b)$$

$$L(P, b) = \exp(\lambda_2(P - b)), \quad (16c)$$

where λ_2 is the negative root of the fundamental quadratic associated with arithmetic Brownian motion,

$$\frac{1}{2}\sigma^2\lambda^2 + \mu\lambda - \rho = 0.$$

Substituting (16c) into (3), we get that $\varepsilon_{\lambda_2}(b) = \lambda_2 b$. Substituting this as well as (16a) and (16b) into (12), we get the optimal barrier

$$b^* = C + \frac{1}{\lambda_2} - \frac{\mu}{\sigma}.$$

Since λ_2 is negative, the optimal abandonment barrier is less than the unit cost C .

4.3 Geometric Brownian motion and concave profit function

Equation (10) clearly shows that the objective function in our abandonment problem is composed of three components, which are the present value of the revenue flow, the present value of the cost flow, and the discount factor. These can be evaluated independently of each other. Therefore variations to the problem structure that perturb any of the three components can be accommodated by re-evaluating the affected component only. This is one of the benefits of the approach taken in this paper. As an example, let P follow geometric Brownian motion as in Section 4.1, but assume that the profit function in our problem is changed from $P_t - C$ to $\sqrt{P_t} - C$. This profit function could result e.g. from a constant elasticity demand curve of the form $D(P) = P^{-0.5}$. A power function of geometric Brownian motion is still a geometric Brownian motion, so that the present value corresponding to (9a) but with P_t replaced by $\sqrt{P_t}$ is readily calculated,

$$W(P) = \frac{\sqrt{P}}{\rho - \frac{1}{2}\mu + \frac{1}{8}\sigma^2}. \quad (17)$$

On the other hand, neither the function $W_0(C)$ nor the Laplace transform $L(P,b)$ will change. The optimal stopping barrier is then calculated through substituting (15b), (15c), and (17) into the markup formula (12). The optimal barrier is

$$b^* = \left[\left(\frac{\beta_2}{\beta_2 - 1/2} \right) \left(\frac{\rho - 1/2\mu + 1/8\sigma^2}{\rho} \right) \right]^2 C^2,$$

while the breakeven price level, i.e. the price yielding zero instantaneous profit, is C^2 .

5 A problem of investment and abandonment

Both the optimal investment problem and the optimal abandonment problem analyzed in the previous sections are examples of one-off termination problems, where the choice variable of interest is a single stopping time. We demonstrate here that the direct approach applies also to regenerative problems with multiple (an infinite sequence of) stopping decisions, and that the first-order condition(s) still have interpretations in terms of elasticities of relevant economic quantities.

The problem that we analyze here is a generalization of both the standard optimal investment problem and the optimal abandonment problem studied in the previous section. This problem has previously been analyzed based on dynamic programming methods by Dixit (1989) (see also Dixit and Pindyck, 1994, pp. 215-218). As before, P is a price that follows a time-homogenous diffusion process, C is a constant unit cost, and the firm has a technology that produces a single unit per unit time. The firm may begin to operate by incurring a fixed lump-sum investment cost I , and may cease to operate by incurring a lump-sum abandonment cost D . Once a firm has ceased to operate, it may begin to operate again by incurring the investment cost I . The value of the firm is the present value of profits less the present value of the costs of investment and abandonment. As a function of the abandonment barrier b_1 and the investment barrier b_2 (b_1 is less than b_2), the value of an idle firm, i.e. one that is not operating, is

$$V(P; b_1, b_2) = E_P \left[\sum_i \int_{\tau_{b_2}^i}^{\tau_{b_1}^i} e^{-\rho t} (P_t - C) dt - \sum_i e^{-\rho \tau_{b_2}^i} I - \sum_i e^{-\rho \tau_{b_1}^i} D \right], \quad (18)$$

where $\tau_{b_1}^1, \tau_{b_1}^2, \dots$ are successive hitting times of the barrier b_1 , and similarly for the hitting times of b_2 . The hitting times are formally defined by

$$\tau_{b_2}^1 = \inf\{t \geq 0 : P_t \geq b_2\},$$

$$\tau_{b_1}^i = \inf\{t > \tau_{b_2}^i : P_t = b_1\}, i = 1, 2, \dots,$$

$$\tau_{b_2}^i = \inf\{t > \tau_{b_1}^{i-1} : P_t = b_2\}, i = 2, 3, \dots$$

This is regenerative problem which starts anew at each stopping time. The firm value can be expressed in a concise form by taking advantage of this regenerative structure, as well as the strong Markov nature of the state variable. Let us define the functions

$$W(P) = E_P \left[\int_0^\infty e^{-\rho t} P_t dt \right] \quad (19a)$$

$$W_0(C) = E_P \left[\int_0^\infty e^{-\rho t} C dt \right] \quad (19b)$$

$$L(b_1, b_2) = E_{b_2} \left[e^{-\rho(\tau_{b_1}^i - \tau_{b_2}^i)} \right] \quad (19c)$$

$$U(b_1, b_2) = E_{b_1} \left[e^{-\rho(\tau_{b_2}^{i+1} - \tau_{b_1}^i)} \right]. \quad (19d)$$

Here $W(P)$ and $W_0(C)$ are exactly as in the previous problem, $L(b_1, b_2)$ is the Laplace transform of the time interval it takes for the state variable P to first-hit b_1 , starting from b_2 , and $U(b_1, b_2)$ is the Laplace transform of the time interval it takes for the state variable P to first-hit b_2 , starting from b_1 . Appendix A shows that, assuming that the firm is idle initially, the objective function (18) can be expressed with the help of (19a-d) as

$$V(P; b_1, b_2) = \left(\frac{U(P, b_2)}{1 - L(b_1, b_2)U(b_1, b_2)} \right) (W(b_2) - W_0(C) - I - L(b_1, b_2)[W(b_1) - W_0(C) + D])$$

We define

$$V^1(P; b_1, b_2) = \left(\frac{U(P, b_2)}{1 - L(b_1, b_2)U(b_1, b_2)} \right),$$

$$V^2(b_1, b_2) = W(b_2) - W_0(C) - I - L(b_1, b_2)[W(b_1) - W_0(C) + D],$$

so that the objective function (18) takes the form

$$V(P; b_1, b_2) = V^1(P; b_1, b_2)V^2(b_1, b_2). \quad (20)$$

Here V^1 is the expected present value of a stream of 1's received at times $\tau_{b_2}^1, \tau_{b_2}^2, \dots$, and is therefore a multiperiod discount factor. V^2 is the expected value of the profit flow during the time interval it takes for the state variable P to first-hit b_1 , given that P starts at b_2 , added with the value of the lump-sum payments at b_1 and b_2 . This is the value of one recurrent sequel in this regenerative problem, the sequels of which are ex ante identical probabilistically. The objective function then has a very natural interpretation as the multiperiod discount factor times the value of a typical sequel. This representation of the objective is entirely analogous to the representations (2) and (10), respectively, in the one-off stopping problems. Therefore it is clear that the first-order conditions will have similar look out as well.

We write the first-order conditions to (20) in terms of the elasticities of the functions V^1 and V^2 with respect to the barriers b_1 and b_2 ,

$$\varepsilon_{V_{b_1}^1}(b_1^*, b_2^*) = \varepsilon_{V_{b_1}^2}(b_1^*, b_2^*), \quad (21a)$$

$$\varepsilon_{V_{b_2}^1}(b_1^*, b_2^*) = \varepsilon_{V_{b_2}^2}(b_1^*, b_2^*), \quad (21b)$$

where the elasticities have been defined as

$$\varepsilon_{V_{b_1}^1}(b_1, b_2) = \frac{\partial V^1(b_1, b_2)}{\partial b_1} \frac{b_1}{V^1(b_1, b_2)}, \quad \varepsilon_{V_{b_1}^2}(b_1, b_2) = -\frac{\partial V^2(b_1, b_2)}{\partial b_1} \frac{b_1}{V^2(b_1, b_2)},$$

$$\varepsilon_{V_{b_2}^1}(b_1, b_2) = -\frac{\partial V^1(b_1, b_2)}{\partial b_2} \frac{b_2}{V^1(b_1, b_2)}, \quad \varepsilon_{V_{b_2}^2}(b_1, b_2) = \frac{\partial V^2(b_1, b_2)}{\partial b_2} \frac{b_2}{V^2(b_1, b_2)}.$$

All elasticities are generally functions of both b_1 and b_2 . In order to interpret the equations, remember that V^1 is a multiperiod discount factor, and V^2 is the value of a single sequel or turnaround in the regenerative problem. As the abandonment barrier b_1 is increased, the expected length of a sequel is reduced, so that the multiperiod discount factor increases.

Raising b_1 at the optimal pair (b_1, b_2) simultaneously reduces the value of a single turnaround, and at the optimum these two effects are balanced according to (21a). As the investment barrier b_2 is increased, the expected length of a sequel is increased, and the multiperiod discount factor is therefore reduced. Raising b_2 at the optimal pair (b_1, b_2) on the other hand increases the value of a single turnaround. At the optimum pair (b_1, b_2) the effects are balanced according to (21b).

The first-order conditions (21a-b) are entirely analogous to the first-order condition (14) in the one-off abandonment problem. The essential difference between (14) and (21a-b) is that in (14) the discount factor is a single period one, whereas in (21a-b) the discount factor is the multiperiod discount factor V^1 .

In order to obtain the value of the objective in closed form as a function of b_1 and b_2 , it is sufficient to know (19a-d) in closed form. When P is a geometric Brownian motion, the expressions for $W(P)$, $W_0(C)$, and $L(b_1, b_2)$ are as in (15a-c). For $U(b_1, b_2)$ we have

$$U(b_1, b_2) = \left(\frac{b_1}{b_2} \right)^{\beta_1},$$

where β_1 is the positive solution to the fundamental quadratic associated with Geometric Brownian motion. We have therefore a closed form expression of the objective as a function of b_1 and b_2 . The first-order conditions for b_1 and b_2 however are quite complicated and it does not appear to be possible to solve them for b_1 and b_2 in closed form.

Dixit (1989) and Dixit and Pindyck (1994) have analyzed this problem based on dynamic programming methods. They also obtained a system of first-order conditions which could not be solved explicitly for the barriers. As compared to our direct approach, the dynamic programming route resulted in a system of four equations in four variables (equations (12-15) in Dixit, 1989). The unknowns in these equations included the two barriers, as well as two unknown coefficients of the general solution to the HJB differential equation. In reference to the solution of Dixit (1989), the direct approach leads us directly into a situation where the unknown coefficients of the HJB equation have been solved in terms of the unknown barriers. Moreover, the first-order conditions in the direct approach do have clear economic interpretations as conditions equating marginal benefits with marginal costs, and have a

format which is familiar from e.g. classical producer theory. This is hardly the case with the first-order conditions (12-15) in Dixit (1989).

Conclusions

We have shown that the direct approach yields first-order conditions with markup interpretations in optimal stopping problems with flow payoffs as well as in regenerative problems. The markup conditions are necessary conditions for optima, while we have not discussed the sufficiency of these conditions. In our specific examples we have demonstrated that the solutions to the markup necessary conditions coincide with those obtained based on dynamic programming arguments. Sufficient conditions for optimal solutions in related problems can be found in the relevant mathematical literature, e.g. Alvarez (2001) and Brekke and Oksendal (1994).

Interesting extensions of the techniques applied here are to problems where the state variable experiences jumps, such as the cash management model of Bar-Ilan et al. (2002). It appears that these problems are quite untractable using dynamic programming methods, but that renewal arguments can be of great value. In general, it appears that the dynamic programming approach has often been pursued in the economic literature in place of the more direct approach, even in cases where the direct approach would have been equally (or more) tractable and conceptually simpler. The models analyzed in this paper, as well as in Sodal (2002), are simple demonstrations of this.

Appendix A. Derivation of a sum-of-geometric-series representation for formula (18)

The objective function (18) is

$$V(P; b_1, b_2) = E_P \left[\sum_i \int_{\tau_{b_2}^i}^{\tau_{b_1}^i} e^{-\rho t} (P_t - C) dt - \sum_i e^{-\rho \tau_{b_2}^i} I - \sum_i e^{-\rho \tau_{b_1}^i} D \right]. \quad (\text{A1})$$

The expectation of the first summation can be written as

$$\begin{aligned}
& E_P \left[\sum_i \int_{\tau_{b_2}^i}^{\tau_{b_1}^i} e^{-\rho t} (P_t - C) dt \right] \\
&= E_P \left[e^{-\rho \tau_{b_2}^1} \left(\int_{\tau_{b_2}^1}^{\tau_{b_1}^1} e^{-\rho(t-\tau_{b_2}^1)} (P_t - C) dt + e^{-\rho(\tau_{b_2}^2 - \tau_{b_2}^1)} \left(\int_{\tau_{b_2}^2}^{\tau_{b_1}^2} e^{-\rho(t-\tau_{b_2}^2)} (P_t - C) dt + \Lambda \right) \right) \right] \quad (A2)
\end{aligned}$$

By the strong Markov nature of the state variable, any of the flow integrals can be written as

$$\begin{aligned}
& E_P \left[\int_{\tau_{b_2}^i}^{\tau_{b_1}^i} e^{-\rho(t-\tau_{b_2}^i)} (P_t - C) dt \right] \\
&= E \left[\int_{\tau_{b_2}^i}^{\infty} e^{-\rho(t-\tau_{b_2}^i)} (P_t - C) dt \right] - E_{b_2} \left[e^{-\rho(\tau_{b_1}^i - \tau_{b_2}^i)} \right] E \left[\int_{\tau_{b_1}^i}^{\infty} e^{-\rho(t-\tau_{b_1}^i)} (P_t - C) dt \right] \quad (A3) \\
&= E_{b_2} \left[\int_0^{\infty} e^{-\rho t} (P_t - C) dt \right] - E_{b_2} \left[e^{-\rho(\tau_{b_1}^i - \tau_{b_2}^i)} \right] E_{b_1} \left[\int_0^{\infty} e^{-\rho t} (P_t - C) dt \right]
\end{aligned}$$

Taking expectations in (A2) term by term, substituting in (A3), and using the notation defined in (19a-d), (A1) becomes

$$U(P, b_2) \left(\frac{W(b_2) - W_0(C) - L(b_1, b_2)(W(b_1) - W_0(C)) + \Lambda}{L(b_1, b_2)U(b_1, b_2)} \right),$$

which the term inside the outer parenthesis is an infinite geometric sum where the multiplier is $L(b_1, b_2)U(b_1, b_2)$ and the initial term is $(W(b_2) - W_0(C) - L(b_1, b_2)(W(b_1) - W_0(C)))$. Hence the first summation in (A1) can be written as

$$\left(\frac{U(P, b_2)}{1 - L(b_1, b_2)U(b_1, b_2)} \right) (W(b_2) - W_0(C) - L(b_1, b_2)(W(b_1) - W_0(C))). \quad (A4)$$

Entirely analogous steps yield the expressions

$$\left(\frac{U(P, b_2)}{1 - L(b_1, b_2)U(b_1, b_2)} \right) I, \quad (A5)$$

$$\left(\frac{U(P, b_2)}{1 - L(b_1, b_2)U(b_1, b_2)} \right) L(b_1, b_2)D, \quad (A6)$$

for the second and the third summations in (A1). Summing (A4), (A5) and (A6) gives the representation found in the text.

References

- Alvarez, L.H.R., 1998, Exit strategies and price uncertainty: a Greenian approach, *Journal of Mathematical Economics*, 29, 43-56.
- Alvarez, L.H.R., 1999, Optimal exit and valuation under demand uncertainty: a real options approach, *European Journal of Operational Research*, 114, 320-329.
- Alvarez, L.H.R., 2001, Singular stochastic control, linear diffusions, and optimal stopping: a class of solvable problems, *SIAM Journal of Control and Optimization*, 39, 1697-1710.
- Bar-Ilan, A., D. Perry, W. Stadje, 2002, A generalized impulse control model of cash management, unpublished manuscript.
- Bensoussan, A., J. L. Lions, 1982, *Impulse control and quasi-variational inequalities*, Gauthier-Villars, Paris.
- Bentolila, S., G. Bertola, 1990, Firing costs and labor demand: how bad is eurosclerosis, *Review of Economic Studies*, 57, 381-402.
- Brekke, K.A., B. Oksendal, 1994, Optimal switching in an economic activity under uncertainty, *SIAM Journal of Control and Optimization*, 32, 1021-1036.
- Dixit, A. K., 1989, Entry and exit decisions under uncertainty, *Journal of Political Economy*, 97, 620-638.
- Dixit, A. K., 1993, *The Art of Smooth Pasting*, Harwood Academic Publishers, Switzerland.
- Dixit, A. K., R. S. Pindyck, 1994, *Investment Under Uncertainty*, Princeton University Press, Princeton, New Jersey.
- Dixit, A. K., R. S. Pindyck, S. Sodal, 1999, A markup interpretation of optimal investment rules, *The Economic Journal*, 109, 179-189.
- Harrison, J.M., 1985, *Brownian Motion and Stochastic Flow Systems*, New York, John Wiley and Sons.
- Leland, H., K. Toft, 1996, Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads, *Journal of Finance*, 51, 987-1019.
- McDonald, R., D. Siegel, 1986, The value of waiting to invest, *Quarterly Journal of Economics*, 101, 707-728.

Oksendal, B., 1995, *Stochastic Differential Equations: An Introduction with Applications*, Springer, Berlin, Heidelberg, New York.

Porteus, E., 1971, On the optimality of generalized (s,S) policies, *Management Science*, 17, 411-426.

Porteus, E., 1972, The optimality of generalized (s,S) policies under uniform demand densities, *Management Science*, 18, 644-646.

Sodal, S., 2002, The stochastic rotation problem: a comment, *Journal of Economic Dynamics and Control*, 26, 509-515.