# Aligning Capital with Risk

A dissertation presented by

Silvan Ebnöther

Supervised by

Prof. Fabio Trojani

Prof. Roberto Ferretti

Prof. Paolo Vanini

Submitted to the

Faculty of Economics

Università della Svizzera italiana

for the degree of

Ph.D. in Economics

September 2015

PhD-Thesis

# Aligning Capital with Risk

Silvan Ebnöther
Università della Svizzera italiana

September 10, 2015

The interaction of capital and risk is of primary interest in the corporate governance of banks as it links operational profitability and strategic risk management. Kaplan and Norton (1992) noted that senior executives understand that their organization's monitoring system strongly affects the behaviour of managers and employees. Typical instruments used by senior executives to focus on strategy are balanced scorecards with objectives for performance and risk management, including an according payroll process. A top-down capital-at-risk concept gives the executive board the desired control of the operative behaviour of all risk takers. It guarantees uniform compensations for business risks taken in any division or business area. The standard theory of cost-of-capital (see e.g. Basel Committee on Banking Supervision (2009)) assumes standardized assets. Return distributions are equally normalized to a one-year risk horizon. It must be noted that risk measurement and management for any individual risk factor has a bottom-up design. The typical risk horizon for trading positions is 10 days, 1 month for treasury positions, 1 year for operational risks and even longer for credit risks. My contribution to the discussion is as follows: In the classical theory, one determines capital requirements and risk measurement using a top-down approach, without specifying market and regulation standards. In my thesis I show how to close the gap between bottom-up risk modelling and top-down capital alignment. I dedicate a separate paper to each risk factor and its application in risk capital management.

| **Risk factor** | **Operational risk** | **Credit risk** | **Market risk** | |
|---|---|---|---|---|
| **Type of business** | Workflows and systems | Lending | Treasury | Trading |
| **Cause of loss** | Failed processes; people and systems | Defaults of counterparties | Interest rate and liquidity risks | Price and volatility risks |
| **Countermeasures** | Checks and balances including separation of competence;<br>project planning and management process | Counterparty validation and rating processes;<br>standardized lending process;<br>credit limit management | Maturity matching;<br>hedge accounting;<br>cash flow modelling | Discipline, continuous hedging with risk constraints |
| **Risk horizon** | Yearly | Multi-year | Monthly | 10 days |
| **Risk-alignment to annual planning process** | In line, i.e. no scaling needed | Down-scaling of long-term risk to one-year | Up-scaling of short term risk to one-year | |
| **Literature about risk measurement** | Little literature on operational risk measurement (and predominant loss data analysis) | Lots of literature about credit risk modelling; no literature on down-scaling | Lots of literature about modelling, pricing and hedging, probability processes and empirical studies | |
| **Contribution of my thesis** | Link between operational workflows and operational risk measurement | Multi-year modelling with down scaling to a one-year horizon and linking to risk capital | Discussion and guidance of up scaling short-term risk measures to a one-year horizon and linking to risk capital | |

Table 1: Risk factors and their different characteristics for risk capital management

The cumulative thesis consists of four individual scholarly papers. Three papers embrace the annual budgeting of risk capital for (i) operational risk, (ii) credit risk, and (iii) market risk. The fourth contribution uses game theory for analysing the allocation of an open line of credit to an entrepreneur.

- Silvan Ebnöther, Paolo Vanini, Alexander McNeil, Pierre Antolinez (2003): Operational Risk: A Practitioner's View, Journal of Risk, 5 (3), pp. 1-16. (NCCR Working Paper No. 52)
- Silvan Ebnöther, Paolo Vanini (2007): Credit portfolios: What defines risk horizons and risk measurement?, Journal of Banking and Finance, 31 (12), pp. 3663-3679. (NCCR Working Paper No. 221)
- Silvan Ebnöther, Markus Leippold, Paolo Vanini (2006): Optimal credit limit management under different information regimes, Journal of Banking and Finance, 30 (2), pp. 463-487. (NCCR Working Paper No. 72)
- Silvan Ebnöther (2015), Economic capital for market risk, Working Paper, Available at Social Science Research Network (papers.ssrn.com)

### Operational risk: A practitioner's view

In June 1999, the Basel Committee on Banking Supervision (”the Committee”) released its consultative document ”The New Basel Capital Accord” (”The Accord”) that proposed a regulatory capital charge to cover ”other risks”. The operational risk is one such ”other risk factor”. Since the publication of the document and its sequels, the industry and the regulatory authorities have been engaged in vigorous and recurring discussions. The banking industry has reached a better understanding and offered a more precise definition of operational risk. By now, the topic of operational risks is dealt with in a more process oriented way. However, in practice, operational risks are subject to a qualitative rather than quantitative approach. Hardly any papers have studied the quantitative modelling of operational risks. The few quantitative papers available propose to collect loss data and to apply extreme value theory on the datasets.

I contribute to these debates from a practitioner’s point of view. To achieve this, I consider a number of issues of operational risk from a case study perspective. The case study is defined for a bank's production unit and factors in self-assessment as well as historical data. The results show that I can define and model operational risk for workflow processes. More specifically, if operational risk is modelled on well-defined objects, all vagueness is dispelled although compared with market or credit risk, a different methodology and different statistical techniques are used. An important insight from a practitioner’s point of view is that not all processes in an organization need to be equally considered for the purpose of accurately defining operational risk exposure. The management of operational risks can focus on key issues; a selection of the relevant processes significantly reduces the costs of defining and designing the workflow items. In a next step, the importance of the four risk factors system failure, theft, fraud and error is analysed using compound Poisson processes and extreme value theory technics. While for quality management all factors matter, fraud and system failure have a non-reliable impact on risk figures. Finally, I am able to link risk measurement to the needs of risk management.

### Credit portfolios: What defines risk horizons and risk measurement?

In this part of the thesis, I describe the setup and calibration of a multi-period credit portfolio model and how to achieve risk figures which are in line with a one-year bank policy.

The strong autocorrelation between economic cycles demands that I analyse credit portfolio risk in a multiperiod setup. I embed a standard one-factor model in such a setup. To be more precise, I work with a synthetic Merton-type one-factor model in which the obligor's future rating grade depends on its synthetic asset value, which is itself a linear combination of one systematic and one idiosyncratic factor. I discuss the calibration of the one-period model to Standard & Poor's ratings data in detail and use a maximum likelihood method as proposed by Gordy and Heidfeld (2002). I extend the model to a multi-period setup by estimating an implicit realization path of the systematic factor. The idea of inversely calculating a realization path was also applied in Belkin et al. (1998). But, I use only historical default rates for our calibration to achieve consistency to the "through-the-cycle" rating definition used by rating agencies. I calibrate a simple time series model to this realization path. The results show a significantly positive autocorrelation. This multi-year extension of the credit portfolio model exactly captures the time dependence structure of the assumed default statistic.

Because single-period risk measures cannot capture the cumulative effects of systematic shocks over several periods, I define an alternative risk measure, which I call the time-conditional expected shortfall (TES), to quantify credit portfolio risk over a multiperiod horizon. This measure extends expected shortfall. Simulating a portfolio similar to the S&P portfolio, I show that using TES as a risk measure, a bank can achieve enough capital cushions to cover losses in credit-risky portfolios if heavy losses in a given year are, as they most probably are, followed by comparable losses in the subsequent years due to the autoregressive behaviour of business cycles.

### Optimal credit limit management under different information regimes

Credit limit management is of paramount importance for successful short-term credit-risk management, even more so when the situation in credit and financial markets is tense. By focusing on limit management, we illuminate an aspect of credit risk modelling different from the traditional approaches initiated by Merton (1974) and Black and Cox (1976). I consider a continuous-time model where the credit provider and the credit taker interact within a game-theoretic framework under different information structures. From a modelling point of view, I assume that the risk and return characteristics of the debtor's investment process affect the bank's limit assessment decision at any given time. The demand for credit following from the optimality of the debtor's investment decision defines an earning component in the bank's value function. In turn, the bank's limit assessment affects the optimization problem of the firm by bounding the possible credit exposure. Therefore, the analysis of credit limit management is defined as the solution of a dynamic non-cooperative game. This setup relates the proposed model to the theory of differential games, first introduced in Isaacs (1954). In particular, the model comes close to the continuous-time model in Holmström and Milgrom (1987), where one agent controls the drift rate vector of a multi-dimensional Brownian motion. However, the presented model differs in the underlying information structure. In addition to incomplete information on the firm's actions, the model features partial information on the state variable. Furthermore, the debtor's credit decision influences both drift and variance of the surplus. The model with complete information provides decision-theoretic insights into the problem of optimal limit policies and motivates more complicated information structures. Moving to a partial information setup, incentive distortions emerge that are not in the bank's interest. I discuss how these distortions can effectively be reduced by an incentive-compatible contract.

### Economic capital for market risk

Market risk results from trading and treasury positions. The effective risk horizons for traders and treasurers are less than one year. It is not reasonable for market risk positions to measure their risk on a one-year horizon due to the implicit assumption that traders or treasurers hold theirs portfolios constant up to the risk horizon. Since they are continuously adapting and hedging their portfolios due to market-shifts, they redeploy their trading positions and treasury exposures several times during a year. Nevertheless, risk capital charges must base on a single enterprise-wide risk horizon that must not coincide with the risk factor specific horizons. The proposed risk consumption approach brings these different views between (i) risk measurement and management on a short-term horizon and (ii) risk capital on a long-term horizon in line.

In this part of the thesis, I present a new approach that I have not encountered in any other scholarly papers. I put forward and analyse an idea of risk budgeting that leads back to discussions with Thomas Domenig, a former risk manager at Zurich Cantonal Bank. A bank assigns an annual risk budget to its market risk managers. The annual risk budget equals the probability of an annual loss higher than a predefined capital-at-risk. Throughout the year, the budget that the risk manager consumes is greater the higher the risk he takes. His risk consumption equates the probability of a year-to-date loss in excess of the capital-at-risk at the 10-day risk horizon. As soon as the risk manager has completely consumed his annual risk budget, he immediately has to hedge his positions.

**References**

Selected literature to cost of capital and balanced scorecard

- Robert S. Kaplan and David P. Norton (1992), The balanced scorecard: measures that drive performance, Harvard Business Review, January-February 1992.
- Franco Modigliani and Merton H. Miller (1958), The cost of capital, corporate finance and the theory of investment, The American Economic Review, 48 (3): 216-297, June 1958.
- Basel Committee on Banking Supervision (2009), Range of practices and issues in economic capital modeling, Bank for International Settlements, Basel.

Operational Risk: A Practitioner's View

- Embrechts, P., C. Klüppelberg and T. Mikosch (1997), Modelling Extremal Events for Insurance and Finance, Springer, Berlin.
- Lindskog, F. and A.J. McNeil (2001), Common Poisson Shock Models, Applications to Insurance and Credit Risk Modelling, Preprint, ETH Zürich.
- Medova, E. (2000), Measuring Risk by Extreme Values, Operational Risk Special Report, Risk, November 2000.

Credit portfolios: What defines risk horizons and risk measurement?

- Allen, L. and Saunders, A.: 2003, A survey of cyclical effects in credit risk measurement models, BIS working papers (126).
- Belkin, B., Forest, L. and Suchower, S.: 1998, A one-parameter representation of credit risk and transition matrices, CreditMetrics Monitor, Third Quarter 1998 pp. 46–56.
- Gagliardini, P. and Gouriéroux, C.: 2005, Stochastic migrations models with application to corporate risk, Journal of Financial Econometrics 3(2), 188–226.
- Gordy, M. and Heitfield, E.: 2002, Estimating default correlations from short panels of credit rating performance data, Federal Reserve Board Working Paper .
- Löffler, G.: 2004, An anatomy of rating through the cycle, Journal of Banking & Finance 28(3), 695–720.
- Merton, R. C.: 1974, On the pricing of corporate debt: The risk structure of interest rates, The Journal of Finance 29(2), 449–70.
- Trück, S. and Rachev, S. T.: 2005, Credit portfolio risk and probability of default confidence sets through the business cycle, The Journal of Credit Risk 1(4).

Optimal credit limit management under different information regimes

- Anderson, R. and Sundaresan, S.: 1996, Design and valuation of debt contract, Review of Financial Studies 9(1), 37–68.
- Black, F. and Cox, J.: 1976, Valuing corporate securities: Some effects of bond indenture provision, Journal of Finance 31(2), 351–367.
- Grossman, S. and Hart, O.: 1983, An analysis of the principal-agent problem, Econometrica 51(1), 7–45.
- Holmström, B. and Milgrom, P.: 1987, Aggregation and linearity in the provision of intertemporal incentives, Econometrica 55, 303–328.
- Isaacs, R.: 1954, Differential games, i, ii, iii, iv, Reports rm-1931, 1399, 1411, 1486, Rand Corporation.
- Kalman, R.: 1960, A new approach to linear filtering and prediction problems, Journal of Basic Engineering 82, 35–45.
- Leland, H.: 1994, Corporate debt value, bond covenants, and optimal capital structure, Journal of Finance 49, 1213–1252.
- Lintner, J.: 1956, Distribution of incomes of corporations among dividends, retained earnings, and taxes, American Economic Review 46, 7113.
- Mirrless, J.: 1971, An exploration in the theory of optimum income, Review of Economic Studies 38, 175–208.

Economic capital for market risk

- H. J. Blommestein, L. H. Hoogduin, and J. J. W. Peeters. Uncertainty and risk management after the great moderation: The role of risk (mis)management by financial institutions. Paper for the 28th SUERF Colloquium on ’The quest for stability’, Utrecht, The Netherlands, 2009.
- Domenig, T., Ebnöther, S. and Vanini, P.: 2005, Aligning capital with risk, Risk Day 2005, Center of Competence Finance in Zurich
- Leo Grepin, Jonathan Tétrault, and Greg Vainberg. After black swans and red ink: How institutional investors can rethink risk management. McKinsey Working Papers on Risk, (17), 2010.
- Grant Kirkpatrick. Corporate governance lessons from the financial crisis. Financial Market Trends, 1(96), 2009.
- OECD. Corporate governance and the financial crisis: Key findings and main massages. June 2009.
- Francesco Saita. Risk Capital Aggregation: The Risk Manager’s Perspective. EFMA 2004 Basel Meetings Paper, 2004.

# Operational Risk: A Practitioner's View

By Silvan Ebnöther[a], Paolo Vanini[b]
Alexander McNeil[c], and Pierre Antolinez[d]

[a] Corporate Risk Control, Zürcher Kantonalbank,
Neue Hard 9, CH-8005 Zurich,
e-mail: silvan.ebnoether@zkb.ch

[b] Corresponding author,
Corporate Risk Control, Zürcher Kantonalbank,
Neue Hard 9, CH-8005 Zurich,
Institute of Finance, University of Southern Switzerland, CH-6900 Lugano,
e-mail: paolo.vanini@zkb.ch

[c] Department of Mathematics,
ETH Zurich, CH-8092 Zurich,
e-mail: alexander.mcneil@math.ehtz.ch

[d] Corporate Risk Control, Zürcher Kantonalbank,
Neue Hard 9, CH-8005 Zurich,
e-mail: pierre.antolinez@zkb.ch

First version: June 2001, this version: October 11, 2002

## Abstract

The Basel Committee on Banking Supervision ("the Committee") released a consultative document that included a regulatory capital charge for operational risk. Since the release of the document, the complexity of the concept of "operational risk" has led to vigorous and recurring discussions. We show that for a production unit of a bank with well-defined workflows operational risk can be unambiguously defined and modelled. The results of this modelling exercise are relevant for the implementation of a risk management framework, and the pertinent risk factors can be identified. We emphasize that only a small share of all workflows make a significant contribution to the resulting VaR. This result is quite robust under stress testing. Since the definition and maintenance of processes is very costly, this last result is of major practical importance. Finally, the approach allows us to distinguish features of quality and risk management respectively.

**Keywords:** Operational Risk, Risk Management, Extreme Value Theory, VaR
**JEL Classification: C19, C69, G18, G21**

*Acknowledgement: We are especially grateful to Professor Embrechts (ETH) for his profound insights and numerous valuable suggestions. We are also grateful to A. Allemann, U. Amberg, R. Hottinger and P. Meier from Zürcher Kantonalbank for providing us with the data and relevant practical insight. We finally thank the participants of the seminars at the Universities of Pavia, Lugano and Zurich and the IBM 2002 Finance Forum, Zurich. Paolo Vanini would like to thank the Swiss National Science Foundation (NCCR FINRISK) for their financial support.*

# 1 Introduction

In June 1999, the Basel Committee on Banking Supervision (”the Committee”) released its consultative document ”The New Basel Capital Accord” (”The Accord”) that included a proposed regulatory capital charge to cover ”other risks”. Operational risk (OR) is one such ”other risk”. From the time of the release of this document and its sequels (BIS (2001)), the industry and the regulatory authorities have been engaged in vigorous and recurring discussions. It is fair to say that at the moment, as far as operational risk is concerned the "Philosopher’s Stone" is yet to be found.

Some of the discussions are on a rather general and abstract level. For example, there is still ongoing debate concerning a general definition of OR. The one adopted by the BIS Risk Management Group (2001) is ”the risk of direct loss resulting form inadequate or failed internal processes, people and systems or from external events.” How to translate the above definition into a capital charge for OR has not yet been fully resolved; see for instance Danielsson et al. (2001). For the moment, legal risk is included in the definition, whereas systemic, strategic and reputational risks are not.

The present paper contributes to these debates from a practitioner’s point of view. To achieve this, we consider a number of issues of operational risk from a case study perspective. The case study is defined for a bank's production unit and factors in self-assessment as well as historical data. We try to answer the following questions quantitatively:

1. Can we define and model OR for the workflow processes of a bank's production unit (production processes)? A production process is roughly a sequence of business activities; a definition is given in the beginning of Section 2.
2. Is a portfolio view feasible and with what assets?
3. Which possible assessment errors matter?
4. Can we model OR such that both the risk exposure and the causes are identified? In other words, not only risk measurement but risk management is the ultimate goal.
5. Which are the crucial risk factors?
6. How important is comprehensiveness? Do all workflows in our data sample significantly contribute to the operational risk of the business unit?

The results show that we can give reasonable answers to all the questions raised above. More specifically, if operational risk is modelled on well-defined objects, all vagueness is dispelled although compared with market or credit risk, a different methodology and different statistical techniques are used. An important insight from a practitioner’s point of view is that not all processes in an organization need to be equally considered for the purpose of accurately defining operational risk exposure. The management of operational risks can focus on key issues; a selection of the relevant processes significantly reduces the costs of defining and designing the workflow items. To achieve this goal, we construct the Risk Selection Curve (RiSC), which singles out the relevant workflows needed to estimate the risk figures. In a next step, the importance of the four risk factors considered is analyzed. As a first result, the importance of the risk factors depends non-linearly on the confidence level used in measuring risk. While for quality management all factors matter, fraud and system failure have a non-reliable impact on risk figures. Finally, with the proposed methodology we are able to link risk measurement to the needs of risk management: For each risk tolerance level of the management there exists an appropriate risk measure. Using this measure RiSC and the risk factor contribution anaylsis select the relevant workflows and risk factors.

The paper is organized as follows. In Section 2 we describe the case study. In Section 3 the results using the data available are discussed and compared for the two models. Further, some important issues raised by the case study are discussed. Section 4 concludes.

# 2 Case Study

The case study was carried out for Zürcher Kantonalbank's Production Unit. The study comprises 103 production processes.

## 2.1 Modelling Operational Risk: Framework

The most important and difficult task in the quantification of operational risk is to find a reasonable model for the business activities[1]. We found it useful, for both practical and theoretical reasons, to think of quantifiable operational risk in terms of directed graphs. Though this approach is not strictly essential in the present paper, for operational risk management full-fledged graph theory is crucial (see Ebnöther et al. (2002) for a theoretical approach). In this paper, the overall risk exposure is considered on an aggregated graph level solely for each process. This approach of considering first an aggregated level is essential from a practical feasibility point of view: Considering the costly nature of analyzing the operational risk of processes quantitatively on a "microscopic level", the important processes have to be selected first.

In summary, each workflow is modelled as a graph consisting of a set of nodes and a set of directed edges. Given this skeleton, we next attach risk information. To this end, we use the following facts: At each node (representing, say, a machine or a person) errors in the processing can occur (see Figure 1 for an example).

*Insert Figure 1 around here.*

The errors have both a cause and an effect on the performance of the process. More precisely, at each node there is a (random) input of information defining the performance. The errors then affect this input to produce a random output performance. The causes at a node are the risk factors, examples being fraud, theft or computer system failure. The primary objective is to model the link between effects and causes. There are, of course, numerous ways in which such a link can be defined. As operational risk management is basically loss management, our prime concern is finding out how causes, through the underlying risk factors, impact losses at individual edges.

We refer to the entire probability distribution associated with a graph as the *operations risk distribution*. In our modelling approach, we distinguish between this distribution and *operational risk distribution*. While the operations risk distribution is defined for all losses, the operational risk distribution considers only losses larger than a given *threshold*.
Operational risk modelling, as defined by the Accord, corresponds to the operations risk distribution in our setup. In practice, this identification is of little value as every bank distinguishes between small and large losses. While small losses are frequent, large losses are very seldom encountered. This implies that banks know a lot about the small losses and their causes but they have no experience with large losses. Hence, typically an efficient organization exists for small losses. The value added of quantitative operational risk management for banks thus lies in the domain of large losses (low intensity, high severity). This is the reason why we differentiate between operations risk and operational risk *if* quantitative modelling is considered. We summarize our definition of operational risk as follows:

**Definition 1** *Quantitative operational risk for a set of production processes are those operations risks which exceed a given threshold value.*

Whether or not we can use graph theory to calculate operational risk critically depends on the existence of *standardized and stable* workflows within the banking firm. The cost of defining

[1] Strictly speaking there are three different objects: Business activities, workflows, which are a first model of these activities, and graphs, which are a second model of business activities based on the workflows. Loosely speaking, graphs are mathematical models of workflows with attributed performance and risk information relevant to the business activities. In the sequel we use business activities and workflows as synonyms.

processes within a bank can be prohibitively large (i) if all processes need to be defined, (ii) if they are defined on a very deep level of aggregation, or (iii) if they are not stable over time.

## 2.2 Data

An important issue in operational risk is data availability. In our study we use both self-assessment and historical data. The former are based on *expert knowledge*. More precisely, the respective process owner valued the risk of each production process. To achieve this goal, standardized forms were used where all entries in the questionnaire were properly defined. The experts had to assess two random events:

1. The frequency of the random time of loss. For example, the occurrence probability of an event for a risk factor could be valued "high/medium/low" by the expert. By definition the "medium" class might, for example, comprise one-yearly events up to four-yearly events.
2. The experts had to estimate maximum and minimum possible losses in their respective processes. The assumed severity distribution derived from the self-assessment is calibrated using the loss history[2]. This procedure is explained in chapter 2.4.

If we use expert data, we usually possess sufficient data to fully specify the risk information. The disadvantage of such data concerns their quality. As Rabin (1998) lucidly demonstrates in his review article, people typically fail to apply the mathematical laws of probability correctly but instead create their own "laws" such as the "law of small numbers". An expert based database thus needs to be designed such that the most important and prominent biases are circumvented and a sensitivity analysis has to be done. We therefore represented probabilistic judgments in the case study unambiguously as a choice among real life situations.

We found three principles especially helpful in our data collection exercise:

1. *Principle I:* Avoid direct probabilistic judgments.
2. *Principle II:* Choose an optimal interplay between experts' know how and modelling. Hence the scope of the self-assessment has to be well defined. Consider for example the severity assessment: A possible malfunction in a process leads to losses in the process under consideration. The same malfunction can also affect other workflows within the bank. Experts have to be awake to whether they adopt a local point of view in their assessment or a more global one. In view of the pitfalls inherent in probabilistic judgments, experts should be given as narrow a scope as possible. They should focus on the simplest estimates, and model builders should perform more complicated relationships based on these estimates.
3. *Principle III:* Implement the right incentives. In order to produce the best result it is important not only to advise the experts on what information they have to deliver, but also to make it clear why it is beneficial for them and the whole institution to do so. A second incentive problem concerns accurate representation. Specifically, pooling behavior should be avoided. By and large, the process experts can be classified in three categories at the beginning of a self-assessment: Those who are satisfied with the functioning of their processes, those who are not satisfied with the status but have so far been unable to improve their performance and, finally, experts who well know that their processes should be redesigned but have no intention of doing so. For the first type, making an accurate representation would not appear to be a problem. The second group might well exaggerate the present status to be worse than it in fact is. The third group has an incentive to mimic the first type. Several measures are possible to avoid such pooling behavior, i.e. having other employees crosscheck the assessment values, and comparing with loss data where available. And ultimately, common sense on the part of the experts' superiors can reduce the extent of misspecified data due to pooling behavior.

[2] The loss history was not used in Ebnöther et al. (2002) because the required details were not available. The soundness of the results has been enhanced by the availability of this extended data.

The historical data are used for calibration of the severity distribution (see Section 2.4). At this stage, we restrict ourselves to noting that information regarding the severity of losses is confined to the minimum/maximum loss value derived from the self-assessment.

## 2.3 The Model

Within the above framework, the following steps summarize our quantitative approach to operational risk:

1. First, data are generated through simulation starting from expert knowledge.
2. To attain more stable results, the distribution for large losses is modelled using extreme value theory.
3. Key risk figures are calculated for the chosen risk measures. We calculate the VaR and the conditional VaR (CVaR)[3].
4. A sensitivity analysis is performed.

Consider a business unit of a bank with a number of production processes. We assume that for workflow i there are 4 relevant risk factors $R_{i,j}$, j = 1,..., 4, leading to a possible process malfunction such as system failure, theft, fraud, or error. Because we do not have any experience with the two additional risk factors external catastrophes and temporary loss of staff, we have not considered them in our model. In the present model we assume that all risk factors are *independent*.

To generate the data, we have to simulate two risk processes: The stochastic time of a loss event occurrence and the stochastic loss amount (the severity) of an event expressed in a given currency. The number $N_{i,j}$ of workflow i malfunctions by risk factor j and the associated severity $W_{i,j}(n)$, n = 1,...$N_{i,j}$, are derived from expert knowledge. $N_{i,j}$ is assumed to be a homogeneous Poisson process. Formally, the inter-arrival times between successive losses are i.i.d, exponentially distributed with finite mean $1/\lambda_{i,j}$. The parameters $\lambda_{i,j}$ are calibrated to the expert knowledge database.

The severity distributions $W_{i,j}$ (n) ~ $F_{i,j}$, for n=1, … , $N_{i,j}$ are estimated in a second step. The distribution of severity $W_{i,j}(n)$ is modeled in two different ways. First, we assume that the severity is a combined Beta and generalized Pareto distribution. In the second model, a lognormal distribution is used to replicate the severity.

If the (i,j)-th loss arrival process $N_{i,j}$ (t), t ≧ 0, is independent from the loss severity process $\{W_{i,j}(n)\}_{n\in N}$ and $W_{i,j}(n)$ has the same distribution for each n and are independent, then the total loss experienced by process i due to risk type j up to time t

$$S_{i,j}(t) = \sum_{n=1}^{N_{i,j}(t)} W_{i,j}(n)$$

is called a compound Poisson process. We always simulate 1 year. For example, 10,000 simulations of S(1) means that we simulate the total first years loss 10,000 times.

The next step is to specify the tail of the loss distribution as we are typically interested in heavy losses in operational risk management. We use extreme value theory to smooth the total loss distribution. This theory allows a categorization of the total loss distribution into different qualitative tail regions[4].

In summary, Model 1 is specified by:

[3] VaR denotes the Value-at-Risk measure and CVaR denotes Conditional Value-at-Risk (CVaR is also called Expected Shortfall or Tail Value-at-Risk (See Tasche (2002)).

[4] We consider the mean excess function $e_1(u) = E[S(1)-u \mid S(1) \geqq u]$ for 1 year, which by our definition of operational risk is a useful measure of risk. The asymptotic behavior of the mean excess function can be captured by the generalized Pareto distribution (GPD) G. The GPD is a two-parameter distribution with distribution function

$$G_{\xi,\sigma}(x) = \begin{cases} 1-(1+\frac{\xi}{\sigma}x)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0, \\ 1-\exp(-\frac{x}{\sigma}) & \text{if } \xi = 0, \end{cases}$$

where $\sigma > 0$ and the support is $[0,\infty)$ when $\xi \geqq 0$ and $[0,-\sigma/\xi]$ for $\xi < 0$. A good data fit is achieved which leads to stable results in the calculation of the conditional Value-at-Risk (see Section 3).

- Production processes which are represented as aggregated, directed graphs consisting of two nodes and a single edge,
- Four independent risk factors,
- A stochastic arrival time of loss events modelled by a homogeneous Poisson process and the severity of losses modeled by a Beta-GPD-mixture distribution. Assuming independence, this yields a compound Poisson model for the aggregated losses.
- It turns out that the generalized Pareto distribution, which is fitted by the POT[5] method, yields an excellent fit to the tail of the aggregate loss distribution.
- The distribution parameters are determined using maximum likelihood estimation techniques.

The generalized Pareto distribution is typically used in extreme value theory. It provides an excellent fit to the simulated data for large losses. Since the focus is not on choosing the most efficient statistical method, we content ourselves with the above choice while being very aware that other statistical procedures might work equally well.

## 2.4 Calibration

Our historical database[6] contains losses that can be allocated to the workflows in the production unit. We use this data to calibrate the severity distribution, noting that the historical data show an expected bias: Due to the relevance of operational risk in the last years, more small losses are seen in 2000 and 2001 than in previous years.

For the calibration of the severity distribution we use our loss history and the assessment of the maximum possible loss per risk factor and workflow. The data are processed in two respects. First, as the assessment of the minimum is not needed since it is used for accounting purposes only, we drop this number. Second, errors may well lead to losses instead of gains. In our database a small number of such gains occur. Since we are interested solely in losses, we do not consider events leading to gains.

Next we observe that the maximum severity assessed by the experts is exceeded in some processes. In our loss history, this effect occurs with an empirical conditional probability of 0.87% per event. In our two models, we factor this effect into the severity value by accepting losses higher than the maximum assessed losses.

Calibration is then performed as follows:

- We first allocate each loss to a risk factor and to a workflow.
- Then we normalize the allocated loss by the maximum assessed loss for its risk factor and workflow.
- Finally we fit our distribution to the generated set of normalized losses. It follows that the lognormal distribution and a mixture of the Beta and generalized Pareto distribution provide the best fits to the empirical data.

In the second simulation, we have to multiply the simulated normalized severity by the maximum assessed loss to generate the loss amount (reversion of the second calibration step).

### 2.4.1 Lognormal Model

In our first model of the severity distribution, we fit a lognormal distribution to the standardized losses.

The lognormal distribution seems to be a good fit for the systematic losses. However, we observe that the probability of occurrence for large losses is greater than the empirical data show.

[5] The Peaks-Over-Threshold (POT) method based on a GPD model allows construction of a tail fit above a certain threshold u; for details of the method, see the papers in Embrechts (2000).

[6] The data range from 1997 to 2002 and contain 285 appropriate entries.

### 2.4.2 Beta-GPD-Mixture Model

We eliminate the drawbacks of the lognormal distribution by searching for a mixture of distributions which satisfies the following properties:

First, the distribution has to reliably approximate the normalized empirical distribution in the domain where the mass of the distribution is concentrated. The flexibility of the Beta distribution is used for fitting in the interval between 0 and the estimated maximum $X_{max}$.
Second, large losses, which probably exceed the maximum of the self-assessment, are captured by the GPD with support the positive real numbers. The GPD distribution is estimated using all historical normalized losses higher than the 90% quantile. In our example, the relevant shape parameter $\xi$ of the GPD fit is nearly zero, i.e. the distribution is medium tailed[7]. To generate the losses, we choose the exponential distribution which corresponds to a GPD with $\xi$=0.
Our Beta-GPD-mixture distribution is defined by a combination of the Beta- and the Exponential distribution. A Beta-GPD-distributed random variable X satisfies the following rules: With probability $\pi$, X is a Beta random variable, and with probability (1-$\pi$), X is a GPD-distributed random variable. Since 0.87% of all historical data exceed the assessed maximum, the weight $\pi$ is chosen such that $P(X > X_{max}) = 0.87\%$ holds.

The calibration procedure reveals an important issue if self-assessment and historical data are considered: Self-assessment data typically need to be processed if they are compared with historical data. This shows that the reliability of the self-assessment data is limited and that by processing this data, consistency between the two different data sets is restored.

## 3 Results

The data set for the application of the above approaches is based on 103 production processes at Zürcher Kantonalbank and self-assessment of the probability and severity of losses for four risk factors (see Section 2.1). The model is calibrated against an internal loss database. Since confidentiality prevents us from presenting real values, the absolute values of *all* results are fictitious but the relative magnitudes are real. The calculations are based on 10,000 simulations. Table 1 shows the results for the Beta-GPD-mixture model.

| | $\alpha = 95\%$ | | $\alpha = 99\%$ | | $\alpha = 99.9\%$ | |
|---|---|---|---|---|---|---|
| | VaR | CVaR | VaR | CVaR | VaR | CVaR |
| Emprirical | 17 | 41 | 60 | 92 | 134 | 161 |
| u = 95%-Quantile | 17 | 55 | 52 | 129 | 167 | 253 |
| u = 97.5%-Quantile | - | - | 59 | 10 | 133 | 165 |
| u = 99%-Quantile | - | - | 60 | 91 | 132 | 163 |
| u = 99.5%-Quantile | - | - | - | - | 134 | 161 |

**Table 1** Simulated data behavior of the tail distribution. "Empirical" denotes the results derived from 10,000 simulations for the Beta-mixture model. The other key figures are generated using the POT[8] model for the respective thresholds u.

Using the lognormal model to generate the severities the $VaR_\alpha$ for $\alpha$ = 95% and 99% respectively are approximately the same. The lognormal distribution is more tailed than the Beta-GPD-mixture distribution that leads to higher key figures for the 99.9% quantile.

[7] The lognormal model bellows to the medium tailed distributions, too. But we observe that the tail behavior of the lognormal distribution converts very slowly to $\xi = 0$. For this reason, we anticipate that the resultant distribution on the yearly total loss will seem to be heavily tailed. Only a large-scale simulation could observe this fact.

[8] From Table 1 and 2 it follows that the POT model yields a reasonable tail fit. For further information on the underlying loss tail behavior and statistical uncertainty of the estimated parameters we refer to Ebnöther (2001).

| | $\alpha = 95\%$ | | $\alpha = 99\%$ | | $\alpha = 99.9\%$ | |
|---|---|---|---|---|---|---|
| | VaR | CVaR | VaR | CVaR | VaR | CVaR |
| Emprirical | 14 | 48 | 55 | 137 | 253 | 512 |
| u = 95%-Quantile | 14 | 68 | 53 | 165 | 234 | 633 |
| u = 97.5%-Quantile | - | - | 53 | 195 | 236 | 706 |
| u = 99%-Quantile | - | - | 55 | 277 | 232 | 911 |
| u = 99.5%-Quantile | - | - | - | - | 252 | 534 |

**Table 2** Simulated data behavior of the tail distribution. Instead of the Beta-mixture model of Table 1 the lognormal model is used for the severity.

We can observe that a robust approximation of the coherent risk measure CVaR is more sensitive to the underlying loss distribution. The Tables also confirm that the lognormal model is more heavily tailed than the Beta-mixture model.

## 3.1 Risk Selection Curve

A relevant question for practitioners is how much each of the processes contributes to the risk exposure. If it turns out that only a fraction of all processes significantly contribute to the risk exposure, then risk management needs only to be defined for these processes.

We therefore analyze how much each single process contributes to the total risk. We consider only VaR in the sequel as a risk measure. To split up the risk into its process components, we compare the risk contributions (RC) of the processes.

Let RC$_\alpha$ (i) be the risk contribution of process i to VaR at the confidence level $\alpha$

$$\mathrm{RC}_\alpha(i) = \mathrm{VaR}_\alpha(P) - \mathrm{VaR}_\alpha(P \setminus \{i\}),$$

where P is the whole set of workflows.

Because the sum over all RC$_\alpha$'s is generally not equal to the VaR, the relative risk contribution (RRC$_\alpha$) (i) of process i is defined as the RC$_\alpha$(i) normalized by the sum over all RC$_\alpha$, i.e.

$$\mathrm{RRC}_\alpha(i) = \frac{\mathrm{RC}_\alpha(i)}{\sum_j \mathrm{RC}_\alpha(j)} = \frac{\mathrm{VaR}_\alpha(P) - \mathrm{VaR}_\alpha(P \setminus \{i\})}{\sum_j \mathrm{RC}_\alpha(j)}.$$

As a further step, for each $\alpha$, we count the number of processes that exceed a relative risk contribution of 1%. We call the resulted curve with parameter $\alpha$, the Risk Selection Curve (RiSC).

*Insert Figure 2 around here.*

Figure 2 shows that on a reasonable confidence level only about 10 percent of all processes contribute to the risk exposure. Therefore only for this small number of processes is it worth developing a full graph theoretical model and analyzing this process in more detail. On lower or even low confidence levels, more processes contribute to the VaR. This indicates that there are a large number of processes of the high frequency/low impact type. These latter processes can be singled out for quality management, whereas processes of the low frequency/high impact type are under the responsibility of risk management. In summary, using RiSC graphs allows a bank to discriminate between quality and risk management in respect of the processes which matter. This reduces costs for both types of management significantly and indeed renders OR management feasible.

We finally note that the shape of the RiSC, i.e. not monotone decreasing, is not a product of modelling.

From a risk management point of view RiSC links the measurement of operational risk to its management as follows: Each parameter value $\alpha$ represents a risk measure and therefore, in Figure 2 on the horizontal axes a family of risk measures is shown. The risk managers possess a risk tolerance that can be expressed with a specific value $\alpha$. Hence, RiSC provides the risk information managers are concerned with.

## 3.2 Risk Factor Contribution

The information concerning the most risky processes is important for splitting the Value at Risk into its risk factors. Therefore we determine the relative risk that a risk factor contributes to the $\text{VaR}_\alpha$ in a similar manner to the former analysis. We define the relative risk factor contribution as

$$\text{RRFC}_\alpha(i) = \frac{\text{VaR}_\alpha - \text{VaR}_\alpha(P \setminus \{i\})}{\sum_{j=1}^{4}(\text{VaR}_\alpha - \text{VaR}_\alpha(P \setminus \{j\}))},$$

with P now the whole set of risk factors.
The resultant graph clearly shows the importance of the risk factors.

*Insert Figure 3 around here.*

Figure 3 shows that the importance of the risk factors is not uniform and in linear proportion to the scale of confidence levels. For low levels, error is the most dominant factor, which again indicates that this domain is best covered by quality management. The higher the confidence level is, the more fraud becomes the dominant factor. The factor theft displays an interesting behavior too: It is the sole factor showing a virtually constant contribution in percentage terms at all confidence levels.

Finally, we note that both results, RiSC and the risk factor contribution, were not known to the experts in the business unit. These clear and neat results contrast with the diffuse and disperse knowledge within the unit about the risk inherent in their business.

## 3.3 Modelling Dependence

In the previous model we assumed the risk factors were independent. Dependence could be introduced though a so-called common shock model (see Bedford and Cooke (2001), Chapter 8, and Lindskog and McNeil (2001)). A natural approach to model dependence is to assume that all losses can be related to a series of underlying and independent shock processes. When a shock occurs, this may cause losses due to several risk factors triggered by that shock.
We did not implement dependence in our case study for the following reasons:

- The occurrence of losses which are caused by fraud, error and theft are independent.
- While we are aware of system failures dependencies, these are not the dominating risk factor. (See figure 3.) Hence, the costs for an assessment and calibration procedure are too large compared to the benefit of such an exercise.

## 3.4 Sensitivity Analysis

We assume that for each workflow and each risk factor the estimated maximum loss is twice the self-assessed value, and then twice that value again. In doing so, we also take into account that the calibration to the newly generated data has to be redone.

| | $\alpha = 95\%$ | | $\alpha = 99\%$ | | $\alpha = 99.9\%$ | |
|---|---|---|---|---|---|---|
| | VaR | CVaR | VaR | CVaR | VaR | CVaR |
| Emprirical | 17 | 41 | 60 | 92 | 134 | 161 |
| Stress scenario 1 | 22 | 45 | 57 | 92 | 129 | 178 |
| Stress scenario 2 | 21 | 48 | 65 | 103 | 149 | 186 |

**Table 3** Stress scenario 1 is a simulation using a maximum twice the self-assessed value. Stress scenario 2 is a simulation using a maximum four times the self-assessed value. A Beta-GPD-mixture distribution is chosen as severity model.

It follows that an overall underestimation of the estimated maximum loss does not have a significant effect on the risk figures since the simulation input is calibrated to the loss history.
Furthermore, the relative risk contributions of the risk factors and processes do not change significantly under these scenarios, i.e. the number of processes which significantly contribute to the VaR remains almost invariant and small compared to all processes.[9]

# 4 Conclusion

The scope of this paper was to show that quantification of operational risk (OR), adapted to the needs of business units, is feasible if data exist and if the modelling problem is seriously considered. This means that the solution of the problem is described with the appropriate tools and not by an ad hoc deployment of methods successfully developed for other risks.

It follows from the results presented that a quantification of OR and OR management must be based on well-defined objects (processes in our case). We do not see any possibility of quantifying OR if such a structure is not in place within a bank. It also follows that not all objects (processes for example) need to be defined; if the most important are selected, the costs of monitoring the objects can be kept at a reasonable level and the results will be sufficiently precise. The self-assessment and historical data used in the present paper proved to be useful: applying a sensitivity analysis, the results appear to be robust. In the derivation of risk figures we assumed that risk tolerance may be non-uniform in the management. Therefore, risk information is parameterized such that the appropriate level of confidence can be chosen.

The models considered in this paper can be extended in various directions. First, if the Poisson models used are not appropriate, they can be replaced by a negative Binomial process (see Ebnöther (2001) for details). Second, production processes are only part of the total workflow processes defining business activities. Hence, other processes need to be modelled and using graph theory a comprehensive risk exposure for a large class of banking activities is derived.

[9] At the 90% quantile, for both stress scenarios the number of “relevant “ workflows (8) remains constant whereas a small reduction from 15 to 14 (13) relevant workflows is observed at the median.

# 5 References

- BIS 2001, Basel Committee on Banking Supervision (2001), Consultative Document, The New Basel Capital Accord, http://www.bis.org.
- BIS, Risk Management Group of the Basel Committee on Banking Supervision (2001), Working Paper on the Regulatory Treatment of Operational Risk, http://www.bis.org.
- Bedford, T. and R Cooke (2001), Probabilistic Risk Analysis, Cambridge University Press, Cambridge.
- Danielsson J., P. Embrechts, C. Goodhart, C. Keating, F. Muenich, O. Renault and H. S. Shin (2001), An Academic Response to Basel II, Special Paper Series, No 130, London School of Economics Financial Markets Group and ESRC Research Center, May 2001
- Ebnöther, S. (2001), Quantitative Aspects of Operational Risk, Diploma Thesis, ETH Zurich.
- Ebnöther, S., M. Leippold and P. Vanini (2002), Modelling Operational Risk and Its Application to Bank's Business Activities, Preprint.
- Embrechts, P. (Ed.) (2000), Extremes and Integrated Risk Management, Risk Books, Risk Waters Group, London
- Embrechts, P., C. Klüppelberg and T. Mikosch (1997), Modelling Extremal Events for Insurance and Finance, Springer, Berlin.
- Lindskog, F. and A.J. McNeil (2001), Common Poisson Shock Models, Applications to Insurance and Credit Risk Modelling, Preprint, ETH Zürich.
- Medova, E. (2000), Measuring Risk by Extreme Values, Operational Risk Special Report, Risk, November 2000.
- Rabin, M. (1998), Psychology and Economics, Journal of Economic Literature, Vol. XXXVI, 11-46, March 1998.
- Tasche, D. (2002), Expected Shortfall and Beyond, Journal of Banking and Finance 26(7), 1523-1537.

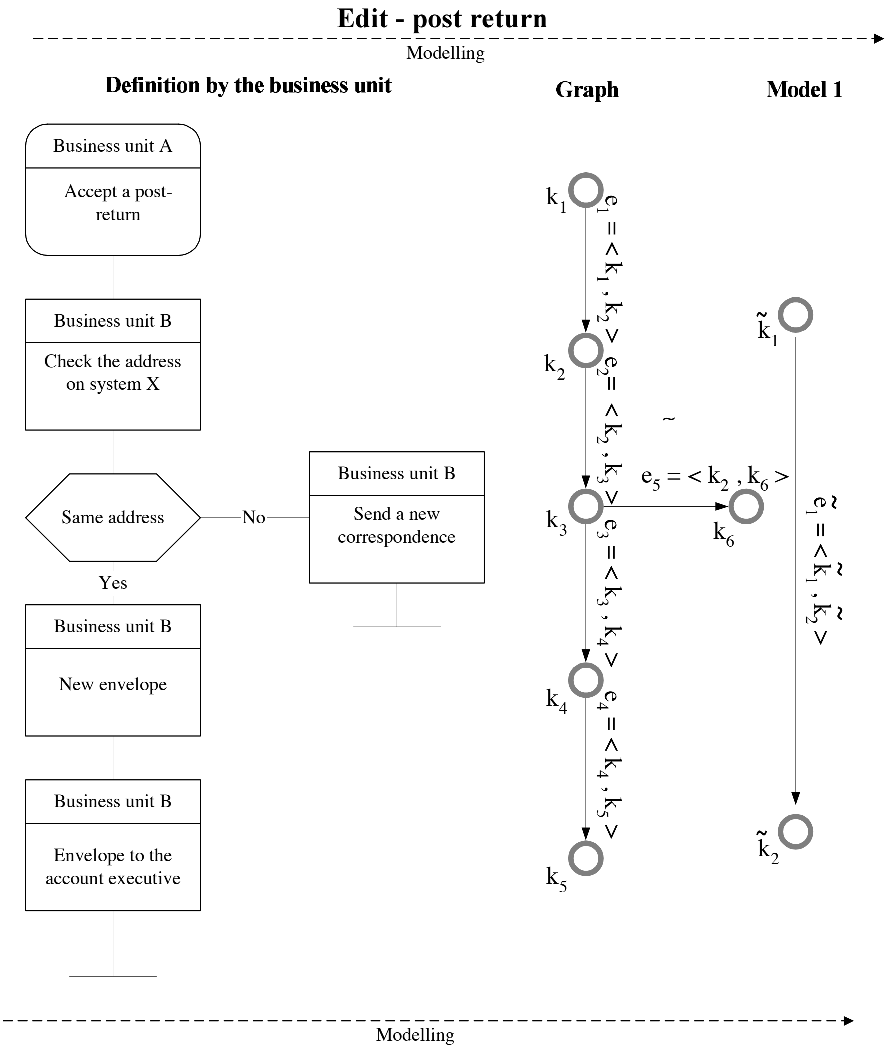


**Figure 1** Example of a simple production process: The edit-post return. More complicated processes can contain several dozens of decision and control nodes. The graphs can also contain loops and vertices with several legs, i.e. topologically the edit - post process is of a particularly simple form. In the present paper, condensed graphs (Model 1) are only considered, while for risk management purposes the full graphs are needed.

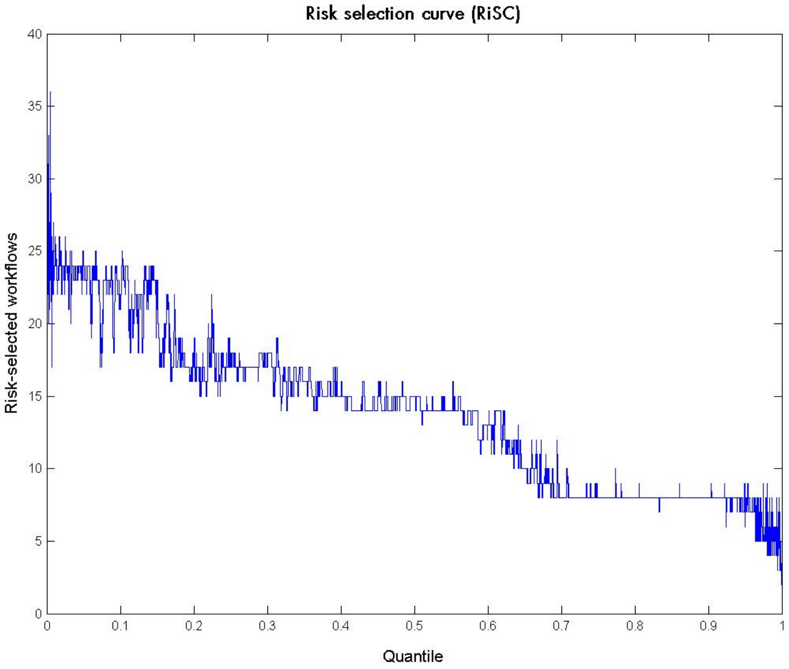


**Figure 2** The risk selection curve (RiSC) of the Beta-GPD-mixture model.

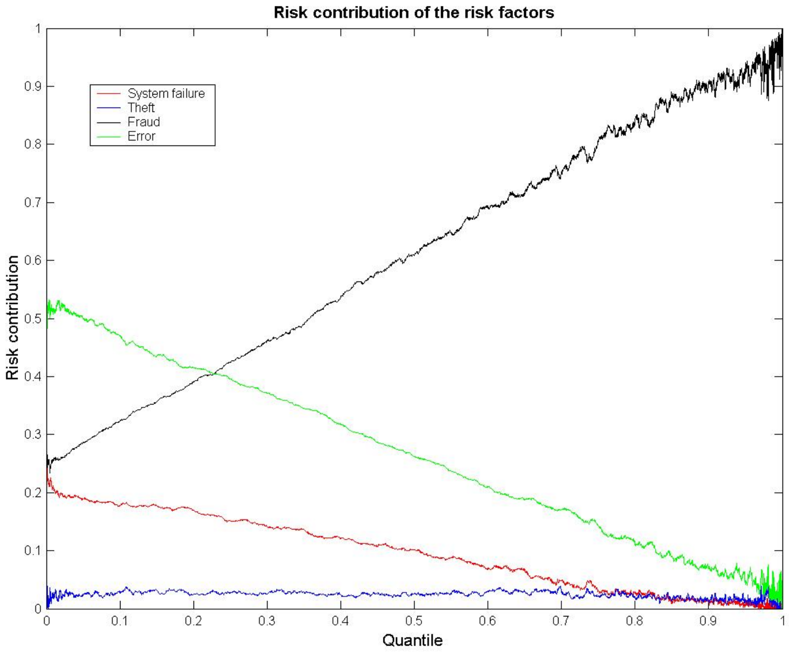


**Figure 3** The segmentation of the VaR into its risk factors.

# Credit portfolios: What defines risk horizons and risk measurement?*

**Silvan Ebnöther**
University of Southern Switzerland, CH-6900 Lugano,
Zürcher Kantonalbank, Josefstrasse 222, CH-8005 Zurich,
e-mail: silvan.ebnoether@zkb.ch

**Paolo Vanini**
Swiss Banking Institute, University of Zürich, CH-8032 Zürich,
Zürcher Kantonalbank, Josefstrasse 222, CH-8005 Zurich,
e-mail: paolo.vanini@zkb.ch

First version: August 2005
This version: Monday 13th November, 2006

**Abstract**

The strong autocorrelation between economic cycles demands that we analyze credit portfolio risk in a multiperiod setup. We embed a standard one-factor model in such a setup. We discuss the calibration of the model to Standard & Poor's ratings data in detail. But because single-period risk measures cannot capture the cumulative effects of systematic shocks over several periods, we define an alternative risk measure, which we call the time-conditional expected shortfall (TES), to quantify credit portfolio risk over a multiperiod horizon.

*JEL Classification Codes: C22, C51, E22, G18, G21, G33.*

**Key words:** Credit risk management, portfolio management, risk measurement, coherence, VaR, expected shortfall, factor model

*Silvan Ebnöther and Paolo Vanini acknowledge the financial support of Zurich Cantonal Bank (ZKB), and Swiss National Science Foundation (NCCR FINRISK) respectively. Corresponding author: Paolo Vanini.

# 1 Introduction

Banks typically measure credit risk over a one-year time horizon, using either value-at-risk (VaR) or expected shortfall (ES) as measures of risk. We claim that the risk horizon has to be longer than one year, because derived risk capital must cover all but the worst possible losses in a credit portfolio. If there is a longer time horizon, then the key risk factor for huge portfolio losses is the economic cycle, which we must model autoregressively. We further extend the standard one-period value-at-risk and shortfall-risk measures to meet the requirements of a multiperiod context. Several facts support these claims. First, although the length, depth, and diffusion of recessions or even depressions has varied significantly in the past, it turns out that a one-year time horizon is often too short to account for a business cycle. For example, when we use the National Bureau of Economic Research (NBER) definition of recessions and depressions, we see that in the U.S. economy for the last 200 years, deep recessions lasted between 35 and 65 months. These long-standing recessions suggest that we should measure credit risk over longer than one year. We believe that a five-year time horizon might be appropriate. We could calculate the risk on a one-year basis and - assuming independence - scale the figures to five years. The reason we do not do so is the autocorrelation in the business cycle: if the industry does badly this year, the probability that it does even worse in the next year is higher than the probability of a strong upwards move. Such an autocorrelation of the business cycle must be accounted for in the risk measurement, else we will underestimate risk in periods of economic downturns. The autocorrelated behavior requires a multiperiod view. The analysis below, which uses Standard & Poor's (S&P) default statistics[1], strongly supports the claim that autocorrelation matters.

Business cycles and factors that are specific to the credit business determine the risk horizon. For example, if we use a risk measure in market risk, then we assume a holding period of ten days with fixed portfolio fractions. There are at least two reasons why such an assumption is useful. First, to try to foresee how portfolios are rebalanced in the future is not realistic. Second, a possible extreme scenario is that trading in a specific period is not, or is almost not, possible. For example, if the liquidity due to a shock event evaporates. Therefore, the risk horizon should also roughly convey during which time changes in the positions are not possible. Two properties define this time for credit risk. First, different types of loan contracts have different maturities and options for exiting and recontracting. Basel II assumes a mean time to maturity of 2.5 years. We can qualitatively confirm this figure if we calculate the mean time of maturity for a portfolio of approximatively 20,000 counterparties. Second, the ability of the institution to buy/sell credit risk on secondary markets is important. The stronger a firm's ability to trade on secondary markets, the shorter the risk's time horizon. These considerations lead us to conclude that that it does not suffice to measure the credit risk of long-term credit investments on a one-year risk horizon. Moreover, based on the significant autocorrelation of default rates, a bank that holds only enough capital to cover one-year

[1] We use public Standard & Poor's data in our case study. The data are described in detail in Standard & Poor's ratings performance 2003, see Standard & Poor's (2004).

losses does not possess enough financial substance to cover multiyear recessions. Therefore, we suggest a credit horizon equal to the maturity of a credit. Since model risk increases with longer risk horizons, it is reasonable to assume a maximal model horizon. We choose a five-year model horizon.

Having established the need for a multiperiod model and a risk horizon longer than a one year, we explain why we need a revision of the usual risk measures. Since we model risk in a multiperiod way, we must define the risk measure for more than one future date. That is, we must show that cumulative losses at different dates in the risk measure lead to meaningful measures of risk. To put it another way, we show that in a multiperiod setup, value-at-risk or expected shortfall underestimate the loss potential in a credit-risky portfolio. We define a new risk measure, which we call time-conditional expected shortfall (TES). TES possesses our required properties. First, it accounts for the fact that heavy credit risk losses can occur in consecutive years. Therefore, TES provides enough capital cushions to survive such events. Second, TES is an extension of expected shortfall and therefore is easily calculated as this one-period risk measure.

The paper is organized as follows. Since our multiperiod model extends a single period one, in Section 2 we reconsider the industry standard one-period, one-factor Merton-type credit risk model. Section 3 extends the model to several periods, i.e., we consider the single economic factor in the Merton model in a multiperiod horizon. Section 4 considers risk measurement in a multiperiod setting. Section 5 concludes.

# 2 One-period model

We work with a synthetic Merton-type one-factor model in which the obligor's future rating grade depends on its synthetic asset value[2], which is itself a linear combination of one systematic and one idiosyncratic factor. Each rating grade is associated with a specified range of obligor's asset value.
Risk management uses factor models for various reasons. First, these models allow us to integrate economic variables. Second, calibration is a standard routine. Third, the models are applicable to internal ratings, since they can be generalized to arbitrary rating definitions and rating grades. Hence, they can be used for small- and medium-sized firms, which represent the majority of counterparties for most institutions. Finally, the Basel Committee also suggests a one-factor model.
Gupton, Finger and Bhatia (1997) and Belkin, Forest and Suchower (1998) show in-depth analysis of one-period factor models and their calibration.

## 2.1 The model

To summarize factor models, we consider a discrete set of ratings $\mathcal{R} = \{1, ..., N_R\}$. We define $N_R$ as the default grade, $R_t^i \in \mathcal{R}$ as the rating grade of obligor $i = 1, ..., N_{obl}$ at

[2] The synthetic asset return is also called distance to default.

time $t$, and $\mathbb{M}_t^i(r,s) = \mathbb{P}\left\{R_t^i = s | R_{t-1}^i = r\right\}$ as the transition probability that obligor $i$ will migrate from rating $r$ to rating $s$ between $t-1$ and $t$.

A single firm defaults as per definition of Merton (1974) if its asset value $A_t^i$ falls below a critical threshold of liability $d^i$. Factor models extend this idea to portfolios of obligors. We use a standard one-factor model to define the portfolio behavior on a one-year risk horizon. The asset value of an obligor is the weighted sum of its joint systematic factor and its idiosyncratic risk factor. That is, the creditworthiness and rating migration probabilities of all obligors depend on one single economic factor, $Z_t$. This key assumption characterizes the dependence structure between the default probabilities of obligors.

**Assumption 2.1.** ***Comparability and dependency***
*Migration probabilities for firms with the same rating grade $r$ are equal. The difference in migration probabilities for each rating grade and scenario depends only on the realization of the market factor $Z$.*

We next formalize the relation between default/migration probabilities and the economic factor $Z$. When we define the synthetic asset return $A_t^i$ of obligor $i$ as a linear combination of the systematic factor $Z_t$ and its idiosyncratic risk $\varepsilon_t^i$ with systematic weight $\omega$, we have:

$$A_t^i := \omega Z_t + \sqrt{1-\omega^2}\,\varepsilon_t^i, \tag{1}$$

where we assume that $Z, \epsilon$ follow a independent standard normal distribution, i.e., $Z_t, \varepsilon_t^i \sim \mathcal{N}(0,1)$ (i.i.d). (We note that we use the shorthand $Z, \varepsilon^i$ for $Z_t, \varepsilon_t^i$ in the one-period setup.) We associate to each rating grade $r$ a range of the asset value. Obligor $i$ with initial rating $r$ at time $t-1$ has rating $s$ at time $t$ if $A_t^i \in [\theta_{r,s+1}, \theta_{r,s})$. The thresholds $\theta_{r,s} \in \mathbb{R}$ satisfy

$$\infty = \theta_{r,1} \geq \theta_{r,2} \geq \ldots \geq \theta_{r,s} \geq \ldots \geq \theta_{r,N_R} \geq \theta_{r,N_R+1} = -\infty. \tag{2}$$

We set $d_r := \theta_{r,N_R}$ for the default thresholds. The thresholds follow from the calibration. Then the migration probabilities of obligor $i$ with initial rating $r$ are

$$\begin{aligned} \mathbb{M}_t^i(r,s) &= \mathbb{P}\left\{R_t^i = s | R_{t-1}^i = r\right\} \\ &= \mathbb{P}\left\{A_t^i \in [\theta_{r,s+1}, \theta_{r,s})\right\}. \end{aligned} \tag{3}$$

Since the synthetic asset returns in equation (1) are still standard normally distributed, the Z-conditional expected default rate of rating grade $r$ reads:

$$p_r(Z) = \Phi\left(\frac{d_r - \omega_r\, Z}{\sqrt{1-\omega_r^2}}\right) \tag{4}$$

with $\Phi(\cdot)$ as the cumulative standard normal distribution function.

## 2.2 Calibration

We calibrate the one-period one-factor model to historical Standard & Poor's data, see Standard & Poor's (2004). To achieve this goal, we must fix the thresholds $\theta_{r,s}$ and the correlation parameters $\omega_r$ for all $r$ and $s$. We use a two-step estimation. First, we fix the threshold values such that the resulting transition probability matrix $\mathbb{M}$ matches the empirical matrix $\overline{\mathbb{M}}$. In the second step, we estimate the risk weights $\omega_r$, for all $r \in \mathcal{R}$.

We calibrate the migration matrix $\mathbb{M}$ is calibrated by dividing the number of firms with rating $s$ at the end of a period by the number of firms with initial rating $r$, i.e.,

$$\overline{\mathbb{M}}_{r,s} = \frac{\sum_t \sharp \text{ Obligors with rating } r \text{ at time } t-1 \text{ and rating } s \text{ at time } t}{\sum_t \sharp \text{ Obligors with rating } s \text{ at time } t-1}. \tag{5}$$

The number of S&P-rated companies increases from 1,371 companies in 1981 to 5,322 companies in 2003. We exclude firms that S&P has not rated (N.R.) at the end of the year. Since there are no registered defaults for the S&P best rating grade AAA, we also exclude AAA firms from our analysis. Furthermore, there is only a single default event for a AA firm. Then the estimated AA default probability $p_2 = 0.01\%$ and the variance $var(p_2) = \sqrt{p_2(1-p_2)/9756} = 0.01\%$ are similar for a sample size of 9,756. Hence, results based on this *AA* statistic have to be adequately considered. The estimated variances for other rating grades are acceptable. The standard normal assumption of the synthetic asset returns implies the calibration condition $\overline{\mathbb{M}}(r,s) = \Phi(\theta_{r,s}) - \Phi(\theta_{r,(s+1)})$. Hence, we fix the default threshold $d_r = \theta_{r,N_r} = \Phi^{-1}(p_r)$ and than calculate the remaining thresholds as

$$\theta_{r,s} = \Phi^{-1}(\overline{\mathbb{M}}(r,s) + \Phi(\theta_{r,(s+1)})) \tag{6}$$

for $s = N_R - 1$ to $s = 2$.

Although the expected loss depends only on exposure and default probabilities, portfolio risk is highly sensitive to the correlation parameter $\omega$. A sophisticated calibration of $\omega$ is more ambitious and different procedures are used in practice, see Gupton et al. (1997), Gordy (2000), Gordy and Heitfield (2002) and Gagliardini and Gouriéroux (2005). A major reason for the difficulty is the short rating-time series. The S&P ratings report, which we use in our calibration, contains a sufficient sample to estimate plausible correlations $\omega_r$ for all $r \in \mathcal{R}$. The easiest and fastest estimation procedure is the moment matching method (MM). We summarize this approach in Appendix A. Gordy and Heitfield (2002) argue that although moment matching methods are common, they often lead to only crude estimations. They propose to use not only some moment statistics, but the whole available data set. A maximum likelihood method promises better estimates, although, due to limited data, this approach can also produce biased parameter estimates.

We use the maximum likelihood method to calibrate the correlations $\omega$. Let $N_r$ and $D_r$ denote the number of obligors and the number of defaults in rating grade $r$. Equation (4) defines the conditional default probability given the market factor $Z$. Using the default probability $p_r := p_r(Z) = p_r(Z|\omega_r)$, the number of defaults $D_r$ in a sample of size $N_r$ follows a binomial distribution with likelihood

$$L(D_r, N_r|p_r) = \binom{N_r}{D_r} p_r^{D_r}(1-p_r)^{N_r-D_r}. \tag{7}$$

The unconditional likelihood of the factor model follows by integrating over the market factor $Z$:

$$\begin{aligned} L(\mathbf{D},\mathbf{N}) &= \mathbb{E}_Z\left[\prod_{r\in\mathcal{R}} L\left(D_r, N_r|p_r(Z|\omega_r)\right)\right] \\ &= \int_{\mathbb{R}} \prod_{r\in\mathcal{R}} \left(\binom{N_r}{D_r} p_r(Z|\omega_r)^{D_r}(1-p_r(Z|\omega_r))^{N_r-D_r}\right) d\Phi(Z). \end{aligned} \tag{8}$$

We next apply equation (8) to the S&P sample from 1981 to 2003. The maximum likelihood estimation $\overline{\omega}$ of vector $\omega$ is

$$\begin{aligned} \overline{\omega} &= \underset{\omega\in[-1,1]^{N_R}}{\arg\max} \prod_{\text{year } t_i} \prod_{r\in\mathcal{R}} L(D_{t_i,r}, N_{t_i,r}|\omega_r) \\ &= \underset{\omega\in[-1,1]^{N_R}}{\arg\max} \prod_{\text{year } t_i} \int_{\mathbb{R}} \prod_{r\in\mathcal{R}} \left(\binom{N_{t_i,r}}{D_{t_i,r}} p_r(Z_{t_i}|\omega_r)^{D_{t_i,r}}(1-p_r(Z_{t_i}|\omega_r))^{N_{t_i,r}-D_{t_i,r}}\right) d\Phi(Z_{t_i}). \end{aligned} \tag{9}$$

Table 1 summarizes the resulting weights $\omega_r$, for all $r \in \mathcal{R}$. Instead of estimating the default probabilities $\mathbf{p} = (p_1, ..., p_{N_r})$ and the systematic weights $\omega = (\omega_1, ..., \omega_{N_r})$ in two separate steps, we also solve the maximization problem (9) in a single step. See Appendix B for the results, which are comparable to those described in the two-step procedure.

| Rating grade | AAA | AA | A | BBB | BB | BB | CCC/C |
|---|---|---|---|---|---|---|---|
| MM | - | 48.75% | 34.57% | 28.94% | 30.84% | 25.83% | 42.97% |
| MLE 1 | - | - | 10.00% | 19.00% | 23.99% | 20.95% | 35.48% |
| MLE 2 | - | - | 19.53% | 18.34% | 22.45% | 18.66% | 23.36% |
| MLE 3 | - | 11.31% | 19.56% | 18.32% | 22.45% | 18.67% | 23.38% |
| Basel II | - | 48.99% | 48.94% | 48.71% | 47.10% | 35.43% | 34.64% |

Table 1: The table shows rating-specific risk weights calibrated to historical default rates of S&P. Since S&P does not register defaults are for AAA companies, we cannot estimate the economic risk factor for such companies. We also doubt the significance for the AA grade, since there is only one default. We use the following abbreviations: MM for moment matching, MLE 1 for five independent estimations for the five weights $\omega_A, ..., \omega_{CCC}$, MLE 2 for one estimation for the five weights, and MLE 3 for one estimation for six weights.

## 2.3 Discussion

We ask if the estimated risk weights in Table 1 are meaningful. The rough intuition that correlations should be between zero and one is fulfilled.[3] For a more refined answer to the question, we compare our estimation with the proposed correlations in Basel II for the internal rating based (IRB) approach:

$$\rho_{\mathrm{IRB}}(p) = 12\% \left( \frac{1 - \exp(-50\, p)}{1 - \exp(-50)} \right) + 24\% \left( 1 - \frac{1 - \exp(-50\, p)}{1 - \exp(-50)} \right). \tag{10}$$

We note that $\omega_{\mathrm{IRB}}^2 := \rho_{\mathrm{IRB}}$. The Basel II correlation depends on obligor characteristics as follows. First, the asset correlation declines with the obligor's probability of default $p$. Second, small- and medium-sized enterprises receive a lower asset correlation, which is not considered in Formula (10). We observe in Table 1 that correlations following Basel II are more conservative than the estimated correlations from S&P ratings data. A possible explanation is that the Basel Committee uses more local or national data compared to the S&P data, which in the last few years has come close to being a global spanning data set. Local economic recessions can intuitively be more pronounced and more frequent than global cycles, during which interference between phase-delayed local cycles can smooth out global economic fluctuations. Hence, we should observe a smaller correlation w for such more geographically diversified portfolios.

A second possible explanation is that Basel II has incentives that differ from those of S&P. The regulator intends a risk monitoring and a capital charge that prevent financial crises. Hence, it is natural that the regulator will add a ”premium” for robustness or model risk for example. Our results are based on historical global data, they are in line with findings in Gagliardini and Gouriéroux (2005) and Gordy and Heitfield (2002).

# 3 Multiperiod modeling

Model extensions to multi-year risk horizons are not standard, neither in the literature nor in practice. A first approach to multiperiod modeling is to apply a one-year model repeatedly, i.e., each year is independently modeled, and the economic latent variable $Z_{t_1}$ is independent of $Z_{t_2}$ for all $t_1 \neq t_2$. This approach conflicts with the time dependence that we observe for both the default rates and the economic fluctuations that affect the creditworthiness of obligors. If time independence can statistically be rejected, an extension of the one-period model needs autocorrelations.

Again, we base the calibration of our multiperiod approach on historical data. First, we invert the one-period model such that we can estimate a realization path $\overline{Z}_t$ of the risk factor $Z_t$ by using a maximum likelihood method. Second, we calibrate a time series model for $Z_t$ to this realization path $\overline{Z}_t$. This multi-year extension of the credit portfolio model exactly captures the time dependence structure of the assumed default statistic.

[3] Full correlation ($\omega = 1$) implies that either all or no company of the portfolio can default. The probability that all companies default is equal to the individual probability of default. On the other hand, independence ($\omega = 0$) implies that an asymptotic portfolio has a constant default rate and cycle effects are not observable in the default rates. Both specifications are evidentially wrong.

## 3.1 Inverting the one-period model

The required data for the model inversion are a representative sample of rated firms and their correlation parameters $\omega_r$, as estimated in Section 2.2. Furthermore, Assumption 2.1 for the one-period model holds. The likelihood $L$ reads as function of the market factor $Z$:

$$\begin{aligned} L(Z_{t_i}) &:= L(Z_{t_i}|\mathbf{D}, \mathbf{N}, \omega) \\ &= \prod_{r\in\mathcal{R}} \left( \binom{N_{t_i,r}}{D_{t_i,r}} p_r(Z_{t_i})^{D_{t_i,r}} (1 - p_r(Z_{t_i}))^{N_{t_i,r}-D_{t_i,r}} \right) \phi(Z_{t_i}). \end{aligned} \tag{11}$$

The inversion procedure maximizes this likelihood function. The solution $\overline{Z_t}$ is a function of $N_{t_i,r}, D_{t_i,r}$ and $\omega_r$. Therefore the maximization problem reads:

$$\begin{aligned} \overline{Z_{t_i}} &= \underset{Z_{t_i}\in\mathbb{R}}{\arg\max}\, L(Z_{t_i}|\omega) \\ &= \underset{Z_{t_i}\in\mathbb{R}}{\arg\max} \prod_{r\in\mathcal{R}} \left( \binom{N_{t_i,r}}{D_{t_i,r}} p_r(Z_{t_i})^{D_{t_i,r}} (1 - p_r(Z_{t_i}))^{N_{t_i,r}-D_{t_i,r}} \right) \phi(Z_{t_i}). \end{aligned} \tag{12}$$

In Equation (12), where we vary $t$ from 1981 to 2003, the path $\overline{Z_{t_i}}$ based on historical S&P data follows. The solution of equation (12) $\overline{Z_{t_i}}$ is based on the default rates of the whole portfolio. We also estimate the implicit factor $\overline{Z}_{t_i,r}$ by the same procedure for each rating grade $r$, with r = AA,..., C.[4] Since the rating-specific samples are smaller, the corresponding results are less robust. Figure 1 shows the realization paths $\overline{Z}_t$ and $\overline{Z}_{t,r}$. The overall factor and the rating-specific factor behave similarly. We consider only the overall realization path model in the sequel.

We discuss the relation of the implicit market factor $\overline{Z}_t$ and economic factors. We choose the gross domestic product (GDP) growth of the U.S. as the economic variable, since the S&P portfolio contains a disproportionate number of U.S. companies. Figure 2 illustrates the synchronic movements of the economy and the default rates. 72.7% of the yearly increments of the two time series have the same sign. The correlation between the implicit market factor $\overline{Z}_t$ and the GDP for the whole time series is 50.0%. In the shorter period, 1985-2003, where we omit the initial years of data collection with only a

[4] The analogous likelihood for each rating grade $r$ is

$$\begin{aligned} L(Z_{t_i,r}) &:= L(Z_{t_i,r}|D_r, N_r, \omega_r) \\ &= \left( \binom{N_{t_i,r}}{D_{t_i,r}} p_r(Z_{t_i,r})^{D_{t_i,r}} (1 - p_r(Z_{t_i,r}))^{N_{t_i,r}-D_{t_i,r}} \right) \phi(Z_{t_i,r}), \end{aligned} \tag{13}$$

and the respective maximization problem is:

$$\begin{aligned} \overline{Z_{t_i,r}} &= \underset{Z_{t_i,r}\in\mathbb{R}}{\arg\max}\, L(Z_{t_i,r}|\omega) \\ &= \underset{Z_{t_i,r}\in\mathbb{R}}{\arg\max} \left( \binom{N_{t_i,r}}{D_{t_i,r}} p_r(Z_{t_i,r})^{D_{t_i,r}} (1 - p_r(Z_{t_i,r}))^{N_{t_i,r}-D_{t_i,r}} \right) \phi(Z_{t_i,r}) \end{aligned} \tag{14}$$

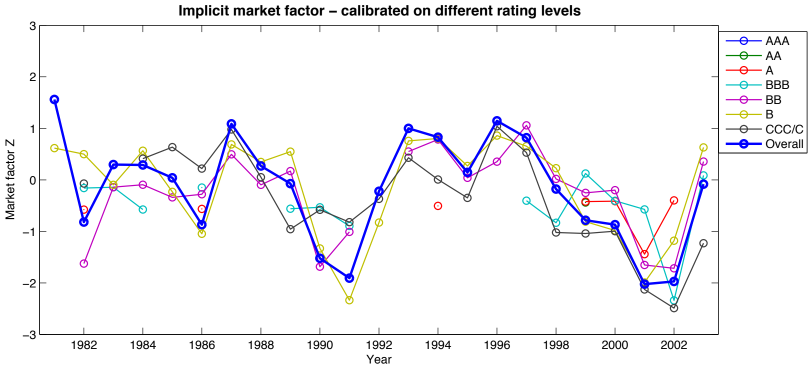


Figure 1: Illustration of the rating-specific implicit market factor $\overline{Z}_{t,r}$ and the overall factor $\overline{Z}_t$. We omit the estimation for years and rating grades in which no defaults occur.

few obligors, correlation increases to 67.3%. Clearly, we can interpret the implicit market factor $\overline{Z_{t_i}}$ based on S&P data as a proxy for the U.S. GDP growth. We conjecture that if we use a multifactor model, the resulting implicit factors would explain GDP growths to a larger extent, and vice versa.

So far, we use historical default rates to calculate the historical implicit economic factor $\overline{Z}_t$. We can apply the same approach to migration rates. We estimate $\overline{Z}_t$ on historical downgrade migration rates and compare default/migration results. We note that we define downgrade migration rates as the migration probability from an initial rating to a lower rating grade, including the default rating grade. Since there are more migration samples, the estimations are more robust. To clarify the notation, we write $Z_t^{def}$ for the implicit path of the market factor calibrated to default rates, and $Z_t^{dwn}$ for the path calibrated to downgrade migration rates. Appendix C summarizes the estimated parameters. Our model works perfectly on an asymptotic portfolio if and only if both observed implicit market factors paths, $\overline{Z}_t^{def}$ and $\overline{Z}_t^{dwn}$, are identical. This is not the case for the S&P sample. Figure 2 shows that perfection is violated in our model: The variation of $\overline{Z}_t^{dwn}$ is typically smaller than the absolute value of $\overline{Z}_t^{def}$.

We claim that the difference is consistent with the rating definition of Standard & Poor's or Moodys. Both rating agencies use long-term ratings to describe the creditworthiness of a company. Allen and Saunders (2003) show that the rating agencies apply "through-the-cycle" rating definition. They use constant stress scenarios which are preferably independent of the current economy. Hence, their ratings should be stable over time. Furthermore, ratings act as more than an order statistic; they do not reflect the actual obligor's default probability in the economic cycle.
On the other hand, rating definitions with constant default probabilities are called

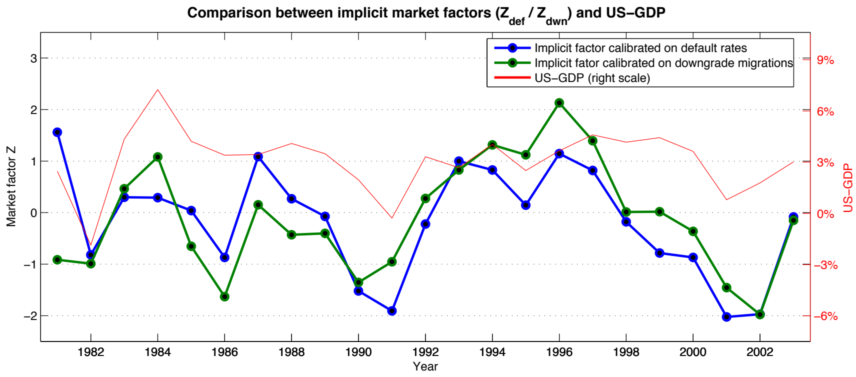


Figure 2: Comparison between implicit market factors $\overline{Z}_t$ calibrated once to default rates and once to downgrade migration rates. The figure also shows the evolution of the U.S.-GDP. We observe a similar behavior of the U.S.-GDP and the estimated implicit factors $\overline{Z}_t$. We identify one long-term boom (1993-1997) and two recessions (1990-1992, 2001-2002) from the historical $Z_t^{dwn}$. Although the GDP is still high in 1998-2000, the implicit factors $\overline{Z}_t^{def}$ and $\overline{Z}_t^{dwn}$ are already decreased, i.e., before the GDP slumped, both indicators had already announced an economic stagnation ($Z_t \sim 0$) followed by a recession ($Z_t \ll 0$).

"point-in-time." Therefore, in a point-in-time model, obligors change their ratings along with the economic cycle, but the vector of default probabilities p is constant through the cycle. See Carey and Hrycay (2003) for an extensive discussion.
Default rates are cyclical, since they are linked to economic cycles,[5] and migration probabilities should remain more or less stable over time by definition. Since the rating agencies use stress scenarios to define the ratings, we expect that the implicit market factors $\overline{Z}_t^{dwn}$ are also more stable over time than is $\overline{Z}_t^{def}$. Our analysis in Figure 2 confirms the expectation: The market factor that we calibrate on the default rates follows the economic cycle, indicated by the U.S. GDP. The implicit factor $\overline{Z}_t^{dwn}$ follows qualitatively the same curve, but is less pronounced than $\overline{Z}_t^{def}$. Hence, the difference between $\overline{Z}_t^{def}$ and $\overline{Z}_t^{dwn}$ is due to the rating definition of Standard & Poor's, and so is not a model misspecification.

We test whether we can reject the time-independence of the market factor based on the two realization paths of the economic variables $\overline{Z}_t^{def}$ and $\overline{Z}_t^{dwn}$. We estimate the correlation parameter, the confidence interval, and the probability (p-value) of the corresponding realization path under the null hypothesis of zero autocorrelation. Table 2 shows the results.

[5] Default rates are above average during recessions.

| underlying data |  |  | estimated | 95%-confidence interval |  |
|---|---|---|---|---|---|
| based on | time | p-value | correlation | lower bound | upper bound |
| default statistic | 1981-2003 | 6.67% | 39.78% | −2.86% | 70.17% |
| default statistic | 1982-2003 | 2.71% | 48.16% | 6.31% | 75.61% |
| downgrade statistic | 1981-2003 | 0.53% | 57.28% | 19.94% | 80.10% |
| downgrade statistic | 1982-2003 | 0.84% | 55.94% | 16.84% | 79.83% |

Table 2: Statistical tests for autocorrelation in the realization paths $\overline{Z}_t^{def}$ and $\overline{Z}_t^{dwn}$.

In the first scenario, which we base on the historical default rates between 1981-2003, we observe that zero autocorrelation cannot be rejected, although the estimated correlation is 39.78%. This statistic is mainly due to the very low default rates in 1981. Figure 3 indicates that the default statistic of 1981 is indeed an outlier. If we exclude the initial year 1981 with only 1,371 companies in the S&P portfolio, autocorrelation increases to 48.16% and the p-value indicates that autocorrelation is statistically significant. For the downgrade migration rates, autocorrelation is highly significant. The p-value is 0.53% and the estimated autocorrelation is 57.28% for the whole time series. The estimated autocorrelation remains stable between 55% and 70% if we stepwise exclude the oldest observations. The high autocorrelation on downgrade rates is also caused by the through-the-cycle rating definition of the rating agencies, since such a definition smoothes out rating changes due to economic cycles.

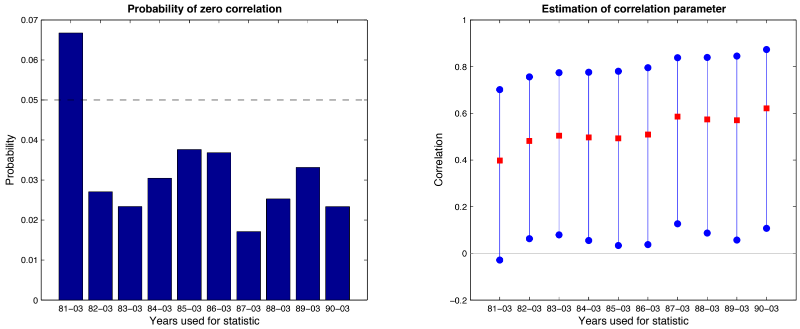


Figure 3: Test for statistical significance to reject zero-autocorrelation based on different time series. The time series differ only in years that we consider in the autocorrelation test.

We also compute a realization path of a mostly Swiss-based credit-risk portfolio with about 18,000 obligors. Since the Swiss economy is less diversified, we expect the economic cycles to be more pronounced than in the global (but U.S. dominated) data of Standard & Poor's. Larger cycles in economy, in the default rates and therefore in the implicit economic factor, are reflected in a higher autocorrelation. Based on historical Swiss-specific default rates, we estimate an autocorrelation of 70% to 80%. Another reason for the slightly higher autocorrelation could be the fact that our Swiss time series is shorter than the S&P data set, i.e., only from 1997 to 2004.

The idea of inversely calculating a realization path $\overline{Z}_t$ was also applied in Belkin et al. (1998). They focused on a sophisticated calibration of a transition matrix in dependence to the economy. Furthermore, they estimated the market factor $Z_t$ using a quadratic minimization of the distance between the empirical migration matrix and a modeled migration matrix which is a function of a realization $Z_t$ in a 1-factor model. We argued above that it is cumbersome to use migration rates to estimate an implicit market factor $Z_t$ due to the the "through-the-cycle" rating definition used by rating agencies: Rating allocations of the rating agencies exclude economic cycles by definition. We intend to achieve a multi-period model for portfolio credit risk measurement. Since possible future defaults define credit risk, we use only historical default rates for our calibration.

## 3.2 The multiperiod extension

Table 2 shows that the economic variable $Z_t$ is statistically autocorrelated. We choose an AR(1) model for the dynamic market factor $Z_t$ to account for the autocorrelation:

$$\begin{aligned} Z_1 &= \xi_1 \\ Z_t &= \beta Z_{t-1} + \sqrt{1-\beta^2}\xi_t, \end{aligned} \tag{15}$$

where all $\xi_t, t = 1, ..., N$, are independent and standard-normal distributed. The autocorrelation weights $\beta$ and $\sqrt{1-\beta^2}$ in equation (15) ensure that the distribution of each $Z_t$ remains standard normal. If we consider the estimation based on migration rates, an autocorrelation $\beta$ between 50% and 65% is plausible. This multiperiod one-factor model is a straightforward extension of the one-period model.

# 4 Risk measurement

We now consider what is perhaps the most important question in portfolio risk management: how can we align measures of portfolio risk with banks' portfolio management objectives? To answer the question we use our new measure of risk, time conditional expected shortfall (TES), which we claim is preferable to the standard risk measures value-at-risk (VaR) and expected shortfall (ES).

The goals of any portfolio risk management are twofold. First, given that risks are appropriately measured, a risk measure should derive a meaningful risk figure such that management can decide whether the level of risk is acceptable or not. Second, given the acceptable risk, the management allocates risk capital to the business units such that there exists a risk-free cushion for the accepted risks, and charges the capital costs to the risk managers. We refer to Artzner, Delbaen, Eber and Heath (1997), Artzner, Delbaen, Eber and Heath (1999) for a general risk measure discussion and to Acerbi and Tasche (2002) for capital allocation. We recall that for a given time horizon $T$, Value-at-Risk $VaR_{99\%}(T)$ only provides a loss threshold that is not exceeded with the probability $\alpha = 99\%$, where ES accounts for the expected excess loss over this threshold. If we apply the multiperiod framework of Section 3, then both the risk measures we describe above must be generalized to several periods. Given the advantages of ES in the single period context, we extend only this measure to several periods. Finally, we use a book-value or actuarial definition of loss. The accounted loss only comprises realized losses in default events. We ignore changes in market value due to other rating migrations.

To motivate TES, we must first define risk measurement appropriately. We claim that two different time horizons matter: T, which is set equal to one year, and H, which is a five-year time horizon. H is the risk horizon that follows from the risk event a bank can survive in all but the worst scenarios. One factor that determines these risk events is the macro economic cycles. In other words, risk is measured on a time horizon such that a bank can survive an economic cycle.

Economic cycles vary in both time and severity, i.e., the impact on economic growth. For example, the data used in the NBER analysis of the last 200 years' recessions and expansions of the U.S. economy show that the average recession lasted for about 20 months. The extremes were around 70 months. The other factor that leads to a time horizon longer than a year is the characteristics of the credit-risky products in the portfolio and the ability of risk management to actively manage credit risk. For example, fixed maturity and interest rate contracts have an average time to maturity that is longer than one year. The calculation in our Swiss-based portfolio gives a value of 2.1 years, which is very close to the suggestions of Basel II in the IRB approach.

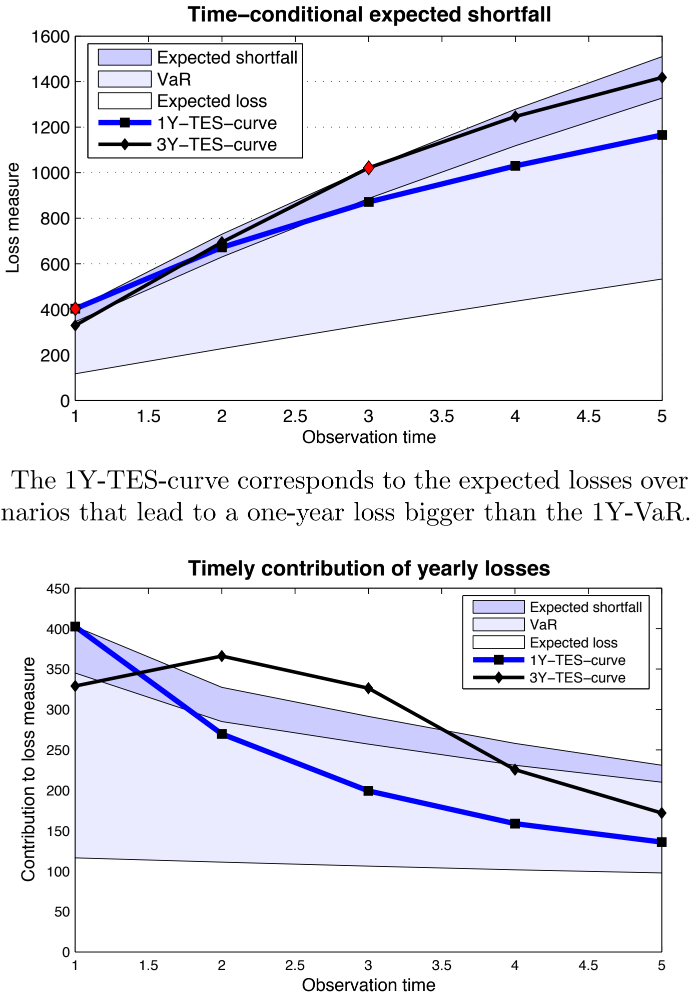
The ability of risk management to manage credit risk is bank specific and can only be roughly estimated. We claim that currently, most institutions are not capable of liquidating positions prior to their maturities. I.e., if creditworthiness deteriorates, the risk management of credit-risky positions is mostly related to reducing unused credit limits, asking for additional collateral, and rasing the price of risk in new contracts. But aside from additional collateral, the bank does not transfer the risk of the actual positions to a secondary market. Given these facts, we choose H to equal five years.

Although we measure risk on a scale of five, business decisions and risk acceptance, i.e., the allocation of risk capital, is made on an annual basis, i.e., $T = 1$ year. This annual basis raises first a technical problem. If $R(q,t)$ is a risk measure on a $t$-time horizon with a creditworthiness goal $q$ - a confidence level, then how do we find $q'$ for $t' \neq t$? In other words, we need to define the relation

$$R(q,T) \leftrightarrows R(q',H) \tag{16}$$

which is necessary to map risk acceptance on a given time horizon consistently to a different time horizon.

Our next step is to determine the risk measure $R$. From a theoretical view, there is only one ”best” answer using dynamic risk measures, see Artzner, Delbaen, Eber, Heath and Ku (2004). There are several reasons why we omit this sound theoretical concept and instead use TES. First, a credit-risk model that is acceptable to the business units,



(a) The 1Y-TES-curve corresponds to the expected losses over the scenarios that lead to a one-year loss bigger than the 1Y-VaR.

(b) Annual loss contributions for different risk measures.

Figure 4: Comparing TES and one-period risk measures VaR and ES. Therefore, we consider a portfolio with properties similar to the S&P portfolio in the year 2004: 5,322 companies, of which 140 possess a rating grade AAA, 497 are AA rated, 1,251 are A rated 1,416 have a BBB, 991 have a BBB, 860 are B, and 167 are CCC/C. We allocate to each company a constant exposure of one unit, hence net exposure at risk is 5,322. We also assume that all credit exposures are hold up to the risk horizon H, which we set equal to five years. We use the correlation parameters $\omega_r$ from the estimations in Section 2.2. Further, we assume an autocorrelation of 60%. We run one million simulations.

and which allows them to calculate individual obligor's risk figures, must be transaction based. Simulating a portfolio based on transactions with approximately $20,000$ obligors generates data sets of teraflop size, i.e., only parallel computers or cluster computing can simulate the risk of such a portfolio on a five-year time horizon. If we use dynamic risk measures with optimal decisions, say on an annual basis, then backward induction starts in year 4, then year 3 is considered, etc. Given the huge number of simulations needed to create a sufficient number of loss statistics, selecting the best dynamic strategy by working backwards is technically not feasible, since the decision rule at each date depends on the value of state variables that are scenario dependent.

A second reason why we propose using our risk measure TES than to work with the fully fledged dynamic approach of Artzner et al. (2004) is the management decision process. Suppose that for today, $t = 0$, we calculate four optimal decisions for the next four years by using dynamic risk measures. Since management allocates capital on an annual basis at $t = 1$, we determine a new four-dimensional optimal strategy vector. This procedure is repeated annually. Therefore, the optimally calculated strategies $t+2, t+3, t+4$ from time $t$ are never actively considered, since they are overwritten by a new set of strategies.

TES is a multiperiod risk measure, but it is not optimal in the dynamic programming sense of Artzner et al. (2004). We claim that the measure is meaningful because it is sensitive to the two time horizons $T$ and $H$, and to risk measurement based on the risk event driven by macro-economic risk. TES consists of two parts. First, it determines shortfall risk on an annual basis, i.e., given the annual risk acceptance by the board of directors, TES calculates mean losses that exceed the VaR. The second term considers risk events that the bank wants to survive over an $H$-time horizon. Therefore, this measures the expected loss difference between time $H$ and $T$, given that risks are not acceptable on the one-year basis. Formally, we have:

$$\begin{aligned}\text{TES}(V(.), H, T) &= \mathbb{E}\left[V(T)|V(T) > VaR_{99\%}(T)\right] + \mathbb{E}\left[V(H) - V(T)|V(T) > VaR_{99\%}(T)\right] \\ &= \mathbb{E}\left[V(T) + \mathbb{E}\left[V(H) - V(T)|V(T) > VaR_{99\%}(T)\right]|V(T) > VaR_{99\%}(T)\right] \\ &= \mathbb{E}\left[V(T) + V(H) - V(T)|V(T) > VaR_{99\%}(T)\right] \\ &= \mathbb{E}\left[V(H)|V(T) > VaR_{99\%}(T)\right] \qquad (17)\end{aligned}$$

**Definition 4.1.** ***Time conditional expected loss (TES):***
*Let $H$ be the risk horizon, $T \leq H$ any date prior to the risk horizon, $V(t)$ the cumulated loss until time $t$ and $VaR_{99\%}(T)$ the 99%-VaR on the cumulative loss at time $T$. We define time-conditional expected shortfall (TES) as:*

$$TES(V(.), H, T) = \mathbb{E}\left[V(H)|V(T) > VaR_{99\%}(T)\right], \qquad H \geq T. \qquad (18)$$

If $T$ and $H$ coincide, then TES reduces to ES. Since at time $T$ a risk constraint is validated, we call $T$ a risk-constraint date and date $H$ the risk horizon.

The risk measure $TES$ reflects the expected cumulative loss at time $H$, conditional that the loss at time $T$ exceeds the chosen quantile (VaR). This measure captures the multiperiod risk of credit portfolios. More precisely, TES measures the conditional expected future trend, given that the annual loss is not acceptable to the bank.

Figure 4 illustrates the following discussion. The figure shows that a year with high losses induces further losses in the following years, i.e., the conditional expected loss increases faster than the average loss. We can observe this accumulation of losses in subsequent years in the data. The historical default rates in the S&P portfolios reaches a peak value in 1990/1991 and in 2000/2001. In both cases, the two consecutive years are in line with the highest default rates. In both cases, an economic risk capital measured by a one-year measure such as VaR or expected shortfall is sufficient to cover the first year losses, but not the cumulative losses over two years.

We can align multiperiod risk calculated using TES risk measure to firm-wide economic capital calculations and allocations that occur on an annual basis. Although firms calculate capital annually using a five-year time horizon, the whole TES value for five years is charged each year to the credit-risk managers. This annual charge reflects the fact that the board of directors approves a credit risk limit each year. In other words, risk managers are annually ”allowed” to lose the equivalent to the calculated TES value. Since the bank's creditworthiness is expressed using a confidence level $q$ on an annual basis $T$, then we must determine the appropriate confidence level $q'$ used for the annual calculation on the five-year time horizon $H$. Since a credit limit is determined each year if the bank has not defaulted, the rule is $q' = q^5$. Using this rule, we assume that the board approves new risk capital independent of the business outcome in the specific years. For details to capital allocation we refer to Domenig, Ebnöther and Vanini (2005).

# 5 Conclusion

Economic cycles - the key credit-risk drivers - show an autoregressive behavior over time. Therefore, if a bank's goal is to survive all but the worst possible events, then its credit-risk events last for several years and must represent this autoregressive behavior. We conclude that credit-risk modeling and measurement requires a multiperiod extension of the traditional one-period models. We calibrate our model to S&P data using historical default rates.

We also define an appropriate multiperiod risk measure. We motivate why a time conditional expected shortfall measure (TES) is well suited to account for an accurate risk measurement in a multiperiod setup. This measure extends expected shortfall. Simulating a portfolio similar to the S&P portfolio, we show that using TES as a risk measure, a bank can achieve enough capital cushions to cover losses in credit-risky portfolios if heavy losses in a given year are, as they most probably are, followed by comparable losses in the subsequent years due to the autoregressive behavior of business cycles.

# A Moment matching

Moment matching methods (MM) are the easiest and fastest way to estimate the systematic risk weights $\omega_r$. To calculate the weights $\omega_r$, we fit the variance of the conditional default probability $var_Z(p_r(Z)) = var_Z(E[\mathbb{1}_{\{A<d_r\}}|Z])$ to the empirical volatility $\overline{\sigma}_r^2$ of the historical default rates. In other words, conditioning w.r.t. $Z$ means averaging over the idiosyncratic effects. The conditional expectation is the corresponding orthogonal

projection on the macro variables. The model variance of the default rate reads:

$$\begin{aligned} var_Z\left(p_r(Z)\right) &= \mathbb{E}_Z\left[p_r(Z)^2\right] - \mathbb{E}_Z\left[p_r(Z)\right]^2 \qquad (19) \\ &= \mathbb{E}_Z\left[\Phi\left(\frac{d_r - \omega_r\, Z}{\sqrt{1-\omega_r^2}}\right)^2\right] - \mathbb{E}_Z\left[\Phi\left(\frac{d_r - \omega_r\, Z}{\sqrt{1-\omega_r^2}}\right)\right]^2 . \end{aligned}$$

We assume $var_Z(p_r(Z)) \equiv \overline{\sigma}_r^2$. This equality leads to an implicit, nonlinear equation which has a unique solution $\omega_r$. We calculate the solution by using the Newton scheme. The results are shown in Table 1.

# B One-step estimation of default probabilities and systematic weights

In Section 2.2, we estimate the default probabilities and systematic weights by applying a two-step procedure. We first calibrate the default probabilities. In a second step, we calibrate the systemic weights. This separation is not necessary, but it leads to more stable results. The concurrent, one-step estimation solves the same maximization problem (9). Hence, we get

$$(\overline{\mathbf{p}}, \overline{\mathbf{w}}) = \underset{\mathbf{p}\in[0,1]^{N_R}, \omega\in[-1,1]^{N_R}}{\arg\max} \prod_{\text{year i}} \prod_{r\in\mathcal{R}} L(D_r^i, N_r^i | \omega_r) \qquad (20)$$

We summarize the estimation of the unconditional default probabilities $\overline{\mathbf{p}}$ and the systematic weights $\overline{\omega}$ in Table 3.

| Rating grade | AAA | AA | A | BBB | BB | BB | CCC/C |
|---|---|---|---|---|---|---|---|
| Mean $p_r$ | 0.00% | 0.01% | 0.05% | 0.39% | 01.49% | 6.76% | 34.80% |
| MLE 1 $p_r$ | - | 0.01% | 0.06% | 0.34% | 01.68% | 6.68% | 29.45% |
| MLE 2 $p_r$ | - | 0.01% | 0.04% | 0.39% | 01.43% | 6.22% | 30.88% |
| MLE 1 $\omega_r$ | - | 0.04% | 12.32% | 19.63% | 25.56% | 20.79% | 32.83% |
| MLE 2 $\omega_r$ | - | $-0.02$% | 13.28% | 20.64% | 24.36% | 18.96% | 24.56% |

Table 3: The table summarizes the default probability. We use the abbreviations as follows: Mean for pure estimate $p$ of as $\widehat{p} = \sum_i(\sharp\text{Defaults in year i}) / \sum_i(\sharp\text{Obligors in year i})$ , MLE 1 for 6 independent estimations of the probabilities $p$ and weights $w$ together, MLE 2 for one estimation for 6 probabilities $p_r$ and 6 weights $w_r$ together.

# C Systematic weights estimated on downgrade migration rates

The calibration procedure of $\overline{Z}_t^{dwn}$ is the same as the presented procedure for $\overline{Z}_t^{def}$, except that we consider historical downgrade migration rates instead of the historical default rates. Hence, we calibrate the "default" threshold $d_r$ to the downgrade probability instead of the default probability. Table 4 shows the estimated weights $\omega$ based on S&P-downgrade statistics.

| Rating grade | AAA | AA | A | BBB | BB | BB | CCC/C |
|---|---|---|---|---|---|---|---|
| D-MM | - | 11.76% | 19.68% | 18.62% | 22.57% | 19.19% | 24.15% |
| D-MLE 1 | 22.25% | 17.08% | 18.42% | 19.94% | 22.65% | 24.84% | 35.39% |
| D-MLE 2 | −1.54% | 12.58% | 15.89% | 14.17% | 15.11% | 20.09% | 22.22% |

Table 4: The table shows rating specific risk weights calibrated to historical downgrade migration rates of S&P. We use the abbreviations: D-MM for moment matching, D-MLE 1 for independent estimations per rating grade, D-MLE 2 for one estimation for all seven weights.

# References

Acerbi, C. and Tasche, D.: 2002, On the coherence of expected shortfall, *Journal of Banking & Finance* **26**, 1487–1503.

Allen, L. and Saunders, A.: 2003, A survey of cyclical effects in credit risk measurement models, *BIS working papers* (126).

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D.: 1997, Thinking coherently, *RISK* **10**(11).

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D.: 1999, Coherent measures of risk, *Mathematical Finance* **9**(3).

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. and Ku, H.: 2004, Coherent multiperiod risk adjusted values and bellmann's principle, *Annals of Operations Research* **forthcoming**.

Bangia, A., Diebold, F. X., Kronimus, A., Schagen, C. and Schuermann, T.: 2002, Ratings migration and the business cycle, with application to credit portfolio stress testing, *Journal of Banking & Finance* **26**(2), 445–474.

Basel Committee on Banking Supervision: 2003, The new basel capital accord, *Consultative document* .

Belkin, B., Forest, L. and Suchower, S.: 1998, A one-parameter representation of credit risk and transition matrices, *CreditMetrics Monitor, Third Quarter 1998* pp. 46–56.

Carey, M. and Hrycay, M.: 2003, Parameterizing credit risk models with rating data, *Journal of Banking & Finance* **25**(1), 197–270.

Crouhy, M., Galai, D. and Mark, R.: 2001, Prototype risk rating system, *Journal of Banking & Finance* **25**, 47–95.

Domenig, T., Ebnöther, S. and Vanini, P.: 2005, Aligning capital with risk, *Risk Day 2005, Center of Competence Finance in Zurich* .

Gagliardini, P. and Gouriéroux, C.: 2005, Stochastic migrations models with application to corporate risk, *Journal of Financial Econometrics* **3**(2), 188–226.

Gonzalez, F., Haas, F., Johannes, R., Persson, M., Toledo, L., Violi, R., Wieland, M. and Zins, C.: 2004, Market dynamics associated with credit ratings a literature review, *European central bank, occasional paper series* (16).

Gordy, M.: 2000, A comparative anatomy of credit risk models, *Journal of Banking & Finance* **24**, 119–149.

Gordy, M.: 2003, A risk-factor model foundation for ratings-based bank capital rules, *Journal of Financial Intermediation* **12**, 199–232.

Gordy, M. and Heitfield, E.: 2002, Estimating default correlations from short panels of credit rating performance data, *Federal Reserve Board Working Paper* .

Gupton, G., Finger, C. and Bhatia, M.: 1997, Creditmetrics, *Technical Document* .

Löffler, G.: 2004, An anatomy of rating through the cycle, *Journal of Banking & Finance* **28**(3), 695–720.

Merton, R. C.: 1974, On the pricing of corporate debt: The risk structure of interest rates, *The Journal of Finance* **29**(2), 449–70.

Nickell, P., Perraudin, W. and Varotto, S.: 2000, Stability of ratings transitions, *Journal of Banking & Finance* **24**, 203–227.

Rösch, D.: 2003, Correlations and business cycles of credit risk: Evidence from bankruptcies in germany, *Financial Markets and Portfolio Management* **17**(3), 309–331.

Schweizerische Nationalbank: 2004, *Statistisches Monatsheft November 2004*, Schweizerische Nationalbank.

Standard & Poor's: 2002, Ratings performance 2001, *Special report* .

Standard & Poor's: 2004, Ratings performance 2003, *Special report* .

Trück, S. and Rachev, S. T.: 2005, Credit portfolio risk and probability of default confidence sets through the business cycle, *The Journal of Credit Risk* **1**(4).

# Optimal Credit Limit Management Under Different Information Regimes[*]

Markus Leippold[†]
*University of Zürich*
Paolo Vanini
*Swiss Banking Institute, University of Zürich and Zürcher Kantonalbank*
Silvan Ebnoether
*Zürcher Kantonalbank*

This Version: February 27, 2005

[*] The authors thank the seminar participants of the Conference "Modeling, Optimization and Risk Management in Finance" at the University of Florida, 2003, and the International Workshop on Risk and Regulation at the Collegium Budapest, 2003. In particular, we would like to thank Giovianni Barone-Adesi, Michel Dacorogna, Hans Geiger, Imre Kondor, Paul Krokhmal, Giorgio Szegö, Ursula Theiler, Stanislav Uryasev, and an anonymous referee for their valuable and thoughtful comments. Markus Leippold and Paolo Vanini acknowledge the financial support of the Swiss National Science Foundation (NCCR FINRISK).

[†] **Correspondence Information:** Markus Leippold, Swiss Banking Institute, University of Zürich, Plattenstrasse 14, 8032 Zürich, Switzerland, tel: +41 1 634 39 62, `mailto:leippold@isb.unizh.ch`

## Abstract

Credit limit management is of paramount importance for successful short-term credit-risk management, even more so when the situation in credit and financial markets is tense. We consider a continuous-time model where the credit provider and the credit taker interact within a game-theoretic framework under different information structures. The model with complete information provides decision-theoretic insights into the problem of optimal limit policies and motivates more complicated information structures. Moving to a partial information setup, incentive distortions emerge that are not in the bank's interest. We discuss how these distortions can effectively be reduced by an incentive-compatible contract. Finally, we provide some practical implications of our theoretical results.

*JEL Classification Codes: G18, C19, G21, C69.*

KEY WORDS: Credit risk management, optimal limit policy, partial information, adverse selection.

In this paper we are concerned with the optimal management of credit limits. For a bank, it is indispensable to have implemented a sound risk management concept, which helps to steer the bank through tense market situations. The drastic stock market downturns in the recent years aggravated the deterioration of the credit markets. As a consequence, the managing of credit risk shifted its focus from a long or mid-term horizon to a very short-term horizon. For credit risk management, short-term policies are mainly concerned with managing credit limits. The credit limit is an approved level of credit allowable that ideally should be compatible with the financial status of the bank's debtor. One of the functions of the bank's credit department is to ensure the proper control of a debtor's account. The credit limit provides a means whereby one aspect of control may be achieved by imposing an upper bound on the potential credit exposure. In the short-term, the management of credit limits is largely the only instrument to control and bound credit losses.

Examples highlighting the importance of proper limit management are numerous. In the aftermath of the LTCM crisis, the United States General Accounting Office noted in their Report to Congressional Requesters of October 1999 that some of LTCM's creditors and counterparties failed to apply appropriate prudential standards. According to the President's Working Group report, such standards include also the management of credit limits on counterparty exposures.

In late 2002, ANZ identified its structured finance division as the major source of the bank's bad debts, as large corporates like Enron collapsed the year before. Most of the losses were related to associated lending rather than the structured finance projects themselves. Consequently, ANZ was "de-risking" their corporate portfolio mostly by lowering credit limits on large single customer exposures. In December 2001, UBS and Credit Suisse offered SWISS, one of many troubled airline companies, a credit limit of CHF500m for restructuring their business. Against the backdrop of the war in Iraq and the spreading of SARS in Asia, outlook for economic growth darkened, in particular for the airline industry. As a result, in April 2003 UBS and Credit Suisse cut their credit limit down to SWISS's actual credit usage, which was by then at CHF100m.

The goal of this paper is to analyze optimal limit management when *a)* the bank has complete information about the company's surplus, *b)* the bank has only partial information about the company's surplus, and *c*) the bank uses an incentive-compatible contract that induces the company to put some effort into revealing information about its true surplus.

The paper is organized as follows. The next section cites the studies that provide background information on credit risk modeling, and defines the position of our work in the literature. Section 2 introduces the notation and presents the model with complete information on the debtor's surplus. In Section 3, we modify the complete information model by assuming only partial information on the debtor's state variable. Since partial information induces some undesirable effects, we introduce in Section 4 an incentive-optimal contract to reduce these effects. The practical implications of the theoretical results are discussed in Section 5. Section 6 concludes.

# 1 Background

In recent years, credit risk management has attracted a lot of attention from the academic community. Three main directions of research in quantitative credit risk management emerged. The first stream analyzes credit portfolios from a diversification point of view. See, e.g., Lucas (1995), Li (2000), Giesecke and Weber (2002), Egloff, Leippold and Vanini (2003), for a recent account. A principal goal is to define reasonable risk and diversification measures and, finally, to determine optimal allocations. A second branch studies the risk transfer of credit positions by either financial or actuarial contracts. Of major concern is the design and valuation of such credit risk contracts. Francis, Frost and Whittaker (1999) and Nelken (1999) provide an exhaustive description of the credit derivative industry. The third stream of literature is focusing on the securitization of credits and loans. See, e.g., Das (2000) among others.

All the above approaches to managing credit risk share a common feature: They can be used to optimize a credit exposure on a mid-term or a long-term basis. Therefore, for all institutions which did not foresee the recent turbulent times in the credit markets, the above instruments were not of much use. E.g., restructuring the balance sheet by transferring credit risk through credit derivatives became almost impossible or, at least, very expensive. When risk cannot be transferred, the focus naturally changes to controlling the current risk exposure on a short-term basis by managing credit limits. During a stock market downturn, the relevance of such a strategy becomes even more accentuated, since heavy losses in stock market values usually lead to pronounced liquidity problems for the bank's credit clients. In many cases, a decrease in the debtor's equity triggers an increase in the demand for credit and loans. Therefore, if banks are not attentive to their credit exposure, it is likely to grow rapidly. By focusing on limit management, we illuminate an aspect of credit risk modelling different from the traditional approaches initiated by Merton (1974) and Black and Cox (1976). In particular, we isolate our analysis from the possibility of default. For the main purpose of this paper, the study of the optimal limit policy under different information regimes, neglecting default is not as severe as it might seem. In practice, the cost of default is often subsumed in the price per unit of credit supplied to the company. In particular, this price is determined by an appropriate margin and several cost components. Taking the perspective of the bank's credit department, these cost components are comprised by *a)* the internal interest rate owed to the treasury department, *b)* the internal production costs, *c)* the regulatory costs, *d)* the costs determined by the internal credit rating. Thus, the company's default probability enters the price per unit of credit supply through the internal credit rating. For a debtor with high default probability, the bank will adjust its price for providing credit accordingly. In our model setup, this price enters the optimization problem of the bank. Therefore, to understand the effect of default, at least in part, we can trace out the corresponding comparative static.

From a modelling point of view, we assume that the risk and return characteristics of the debtor's investment process affect the bank's limit assessment decision at any given time. The demand for credit following from the optimality of the debtor's investment decision defines an earning component in the bank's value function. In turn, the bank's limit assessment affects the optimization problem of the firm by bounding the possible credit exposure. Therefore, the analysis of credit limit management is defined as the solution of a dynamic non-cooperative game. This setup relates our model to the theory of differential games, first introduced in Isaacs (1954). In particular, our model comes close to the continuous-time model in Holmström and Milgrom (1987), where one agent controls the drift rate vector of a multi-dimensional Brownian motion. However, the model that we present differs in the underlying information structure. In addition to incomplete information on the firm's actions, our model features partial information on the state variable. Furthermore, the debtor's credit decision influences both drift and variance of the surplus.

## 2 Limit Policy with Complete Information

In this section, we consider a model with complete information.[1] We define a financial market with a terminal time $T$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Uncertainty is modelled by $W_t$, a one dimensional Brownian motion. Our economy is populated by a company $C$ representing the credit demand side. The credit supply is provided by a bank $B$. We assume that both players $B$ and $C$ are rational and maximize expected utility over a period $[t, T]$, $0 \leq t < T$, with $T$ possibly equal to infinity.

The company $C$ maximizes expected surplus given an upper limit for the surplus' variance. The surplus, denoted by $S_t$, is defined as the difference between assets and liabilities and serves as the state variable in our model. The company (debtor) chooses the optimal credit amount to maximize the expected surplus. This credit demand is expressed in terms of a fraction $c_t$ of the current surplus $S_t$, i.e., the credit demand in absolute terms equals $c_t S_t$.

[1] In practice, lombard loan relationships can be casted into such a model setup.

The bank, in turn, chooses the optimal limit policy $\ell_t$ to maximize earnings minus costs from credit lending. Hence, in our model, the choice variables are *a)* the credit demand $c_t$ as a fraction of the surplus $S_t$, and *b)* the credit limit $\ell_t$ provided by the bank.

If the company does not demand any credit, the surplus is assumed to evolve as

$$dS_t = \mu_0 dt + \sigma_0 dW_t.$$

As soon as the company demands credit, we assume that these additional resources are invested in a project with different return dynamics, such that the stochastic differential equation (SDE) for the surplus changes to

$$dS_t = (\mu_0 + (\mu_1 - p)c_t)\, dt + (\sigma_0 + \sigma_1 c_t)\, dW_t.$$

The fraction of surplus $c_t$ affects the future surplus two-fold. First, it changes the drift of the surplus. The money borrowed by the firm is invested to obtain a return on the new investment, $\mu_1$, that may differ from the actual one, $\mu_0$. From the new return on investment we subtract the price $p > 0$ paid for one unit of the loan. Second, investing in new projects also alters the risk of the present business, which is reflected by the additional volatility parameter $\sigma_1$. Therefore, by lending money, the company can increase the mean of the surplus, but at the same time the company increases its surplus volatility.

We denote by $J^C(S,t)$ the value function of company $C$ and assume that $C$ solves the following optimization:

$$(\mathbf{C}') \quad : \quad \left\{ \begin{array}{lll} J^C(S,t) & = & \max_{c_t} \mathbb{E}\left[ \int_t^T e^{-\delta(s-t)} S_s ds \,|\mathcal{F}_t \right], \\ & s.t. & \int_t^T \mathrm{Var}\left[ e^{-\delta(s-t)} S_s \,|\mathcal{F}_t \right] ds \leq \varsigma^2, \\ & & dS_t = (\mu_0 + (\mu_1 - p)c_t)\, dt + (\sigma_0 + \sigma_1 c_t)\, dW_t, \\ & & c_t S_t \leq \ell_t \ , \quad 0 \leq c_t, \ \forall t \in [0,T]. \end{array} \right.$$

We henceforth abbreviate the company's set of constraints by $\mathcal{C}$. The optimization problem $(\mathbf{C}')$ mimics the single-period mean-variance approach of Markowitz (1952). However, contrary to Markowitz's static model, we formulate the optimization problem in continuous time. Furthermore, instead of maximizing the end-of-period surplus subject to an end-of-period variance bound, we assume that the company $C$ is concerned with the regularity and smoothness of surplus evolution. Indeed, most firms have an incentive to smooth the variation of the surplus. If surplus is too erratic over time, firms face difficulties in explaining their course of business to investors, the public, and other stakeholders, and finally to the banks. Furthermore, smoothing may result from a tax minimizing strategy as originally documented in Lintner (1956). Therefore, the firm maximizes the expected surplus rate over the time horizon $T - t$ given a variance bound. Thus, the risk has to be smaller than a given acceptance level along the whole optimal investment path and not only at the investment horizon. In addition to the variance constraint, the inequality constraint $c_t S_t \leq \ell_t$ puts an upper bound on the credit demand $c_t$ expressed as a fraction of the current surplus. This amount cannot be larger than the amount of the credit limit $\ell_t$. Finally, we assume that $c_t \geq 0$.

Independent of the bank's decision problem, the optimization problem of company $C$ is difficult to solve in the form $(\mathbf{C}')$, since it is not separable in the dynamic programming sense. The non-separability can be circumvented as follows: Independent of the dynamics of the surplus and other linear constraints, the company's preferences in $(\mathbf{C}')$ are equivalent to

$$J^C(S,t,\omega) \;=\; \max_{c_t \in \mathcal{C}} \left( \int_t^T \mathbb{E}\left[ e^{-\delta(s-t)} S_s \,|\mathcal{F}_t \right] - \omega \mathrm{Var}\left[ e^{-\delta(s-t)} S_s \,|\mathcal{F}_t \right] ds \right) \quad (1)$$

with $\mathcal{C}$ the same feasible set as in $(\mathbf{C}')$. The parameter $\omega$ reflects the trade-off between return and risk. Now, consider the separable problem $(\mathbf{C1})$, which is a two-parameter model $\omega, \lambda$:

$$(\mathbf{C1}): \quad J^C(S,t,\omega,\lambda) \;=\; \max_{c_t \in \mathcal{C}} \mathbb{E}\left[ \int_t^T e^{-\delta(s-t)} \left( \lambda_s S_s - \omega S_s^2 \right) ds \,|\mathcal{F}_t \right].$$

Then, the following result holds:[2]

**Proposition 1.** *If $c^*$ is a solution to problem $(\mathbf{C}')$, and if the surplus follows a linear SDE of the form,*

$$dS_t = (a_t + b_t S_t)dt + (\alpha_t + \beta_t S_t)dW_t, \quad a_t + \alpha_t \beta_t \geq 0,$$

*then it is also a solution to problem $(\mathbf{C1})$ with $\lambda_t^* = 1 + 2\omega c_t^*$, where $S^*$ is the wealth trajectory corresponding to $c^*$.*

We first remark that the inequality part of the sufficient condition in Proposition 1 is not as restrictive as it might seem. In our present model, it is reasonable to assume positive values for the parameters $a_t$ and $\alpha_t$. Then, if $\beta_t$ obtains a negative value, this would mean that a higher surplus and an increase in credit usage would reduce the instantaneous surplus variance. Since such a setup is counterintuitive, we can safely assume that $a_t + \alpha_t \beta_t \geq 0$ holds.

Having presented the company's $C$ setup, we next consider the banks de-

[2] For a proof, we refer to Zhou and Li (2000) and Leippold, Trojani and Vanini (2004) in a similar setting.

cision problem (**B**). It reads

$$
(\mathbf{B1}) \quad : \quad \begin{cases} J^B(S,t) &= \max_{\ell_t} \mathbb{E}\left[ \int\limits_t^T e^{-\delta(T-s)} \left(pc_s S_s - \kappa \ell_s\right) ds \,|\mathcal{F}_t \right], \\ s.t. & \ell_t \geq 0, \\ & dS_t = (\mu_0 + (\mu_1 - p)c_t)\, dt + (\sigma_0 + \sigma_1 c_t)\, dW_t. \end{cases}
$$

Following (**B1**), the bank chooses the limit amount $\ell_t$ over the duration $T-t$ such that expected earnings from lending money to company $C$ minus the capital costs $\kappa$ of providing limit $\ell_t$ are maximized. The costs of economic capital are calculated on the limit amount set off for the client and not on the actual credit exposure. If payment were on the credit exposure, the decision problem would become trivial: The bank would fulfill any credit request of the company $C$ as long as $p > \kappa$. The assumption of a constant $\kappa$ reflects the situation of a bank applying the Basel II's standard approach. If the bank would use an internal model approach, $\kappa$ would not be constant over different limit amounts. However, for our analysis we keep $\kappa$ constant.

Since the decision variables of each player affects the value function of the other one and the variables are functions of time, the problems (**C1**) and (**B1**) define a non-cooperative game. To obtain a solution we apply the concept of a subgame-perfect Nash equilibrium: A pair $(c^*, \ell^*)$ is a Nash equilibrium, if

$$
J^B(S,t,c^*,\ell^*) \geq J^B(S,t,c^*,\ell) \ , \ J^C(S,t,c^*,\ell^*) \geq J^C(S,t,c,\ell^*)
$$

for any feasible policies $c$ and $\ell$ and where $c^*$ $(\ell^*)$ is the optimal strategy of the company (bank).

Before trying to solve the game defined above, the following argument simplifies the analytics essentially. We claim that for the equilibrium strategy $c_t^* = \frac{\ell_t^*}{S_t}$ holds. To see this, we denote by $v_t$ the bank's instantaneous utility at time $t$. Suppose that for $\epsilon > 0$, $c_t^* S_t^* + \epsilon = \ell_t^*$ is a Nash equilibrium. Then,

$$
v_t^\epsilon \;=\; pc_t S_t - \kappa \ell_t = p(\ell_t^* - \epsilon) - \kappa \ell_t = (p-\kappa)\ell_t - p\epsilon = v_t - p\epsilon \ . \tag{2}
$$

Therefore, the bank is always better off choosing a limit policy with $\epsilon = 0$ instead of $\epsilon > 0$. In other words, we have only to solve one optimization problem. The optimality for the other player then follows at once.

**Proposition 2.** *Consider an economy with two players solving* (**C1**) *and* (**B1**)*, respectively. The strategies*

$$
c_t^* \;=\; \frac{\ell_t^*}{S_t} \ , \ \ell_t^* = \gamma_1 S_t + \gamma_2 S_t^2,
$$

*are a subgame perfect Nash equilibrium, where*

$$\gamma_1 = \frac{\sqrt{\delta}}{\sigma_1}, \ \gamma_2 = \frac{\mu_0(\mu_1 - p) - \delta\sigma_0\sigma_1}{\sqrt{\delta}\sigma_1\left(\mu_1 - p + \sqrt{\delta}\sigma_1\right)}.$$

From Proposition 2 it follows that the value functions of the bank and the company in equilibrium are quadratic functions in $S_t$.

**Proposition 3.** *Consider an economy with two players solving* (**C1**) *and* (**B1**)*, respectively. The value function of the bank reads*

$$J^B(S,t) = e^{-\delta(T-t)}(p-\kappa)\left(b_0 + b_1 S_t + \frac{1}{2}b_2 S_t^2\right), \qquad (3)$$

*and the value function of the company is given by*

$$J^C(S,t) = e^{-\delta(T-t)}\left(k_0 + k_1 S_t + \frac{1}{2}k_2 S_t^2\right), \qquad (4)$$

*where the constants* $b_i, k_i, i = 1,2,3,$ *are given in the Appendix, equations (A.1) to (A.3), and (A.4) to (A.6), respectively.*

From Proposition 2 we further note that the optimal credit demand $c_t^*$ is an affine function of the surplus $S_t$ with a marginal derivative of $\gamma_2$. It turns out that the sign of this derivative characterizes the equilibrium solution of the two player economy. The proposition also shows that it may be optimal for both the bank and the company to terminate their credit relation even if the surplus is positive. This is the case when the value of $S$ equals $S = -\gamma_1/\gamma_2$. This ratio can only become positive if $\gamma_2 < 0$. With $\gamma_2 < 0$, the optimal limit policy $\ell^*$ is concave in $S$. The sign of $\gamma_2$ not only determines whether the optimal limit policy is either convex or concave in $S$, but also determines the statistical properties of the surplus dynamics. More precisely, it determines whether $S$ follows a stationary or a non-stationary process. Using the optimal policies in the surplus dynamics, the surplus is stationary if and only if

$$(\mu_1 - p)\gamma_2 > 0.$$

Figure 1 plots the optimal limit $\ell^*$ as a function of the company's surplus $S$. When the riskiness of the company's investments is relatively small compared to the corresponding returns, we get $\gamma_2 > 0$ and the optimal limit is a convex function of the surplus. For $\gamma_2 = 0$, the optimal limit is just a straight line with slope $\gamma_1$. For $\gamma_2 < 0$, the optimal limit as a function of surplus becomes concave. In Panel (A) of Figure 2, we plot a possible trajectory of the surplus. We consider the cases $\gamma_2 > 0$ (dotted line, company 1) and $\gamma_2 < 0$ (solid line).

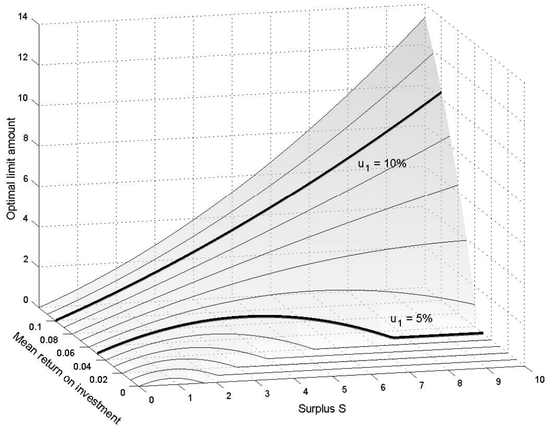


Figure 1: Optimal limit policy. We make the following assumptions: $\mu_0 = 2\%$, $\sigma_0 = 10\%$, $\sigma_1 = 30\%$, $\delta = 5\%$, $p = 1\%$. In order to generate optimal limit policies where $\gamma_2$ switches sign, we set $0 < \mu_1 < 12\%$. As examples, we highlight the optimal policy for $\gamma_2 > 0$ by assuming $\mu_1 = 10\%$, and the optimal policy for $\gamma_2 < 0$ by assuming $\mu_1 = 5\%$.

As we see from Panel (A), the trajectories for $S_t$ when $\gamma_2 > 0$ and $\gamma_2 < 0$ are almost indistinguishable, at least for the first five years. However, the stationarity property has a strong effect on the optimal limit policy of the bank. In Panel (A), we add the relative credit demands for the two cases $\gamma_2 > 0$ and $\gamma_2 < 0$. Comparing the growth of the credit demand and the surplus that may serve as an indicator of the company's health, we see that the company demands credit in a procyclical manner if $\gamma_2 > 0$ and anticyclical if $\gamma_2 < 0$. Therefore, when the surplus is decreasing, a company that has not the opportunity to invest in a project with high expected return, i.e., $\mu_1$ is small, reduces its relative credit demand. Furthermore, since the bank decreases the credit limits, the company is locked out of new investment. On the other side, a company that can invest in highly profitable projects, i.e., $\mu_1$ is large, tends to invest more in relative terms when the economy goes down.

Panel (A) of Figure 2 plots the paths of the optimal limit policies and their corresponding surplus when $\gamma_2 > 0$ (dotted line) and $\gamma_2 < 0$ (solid line). When the surplus is a non-stationary process, the optimal limit process lies considerably above the optimal limit when the surplus process is stationary. In Panel (B), we also plot the bank's profit from providing credit limits. Since with $\gamma_2 > 0$ the company has an increased demand for credit limits, the bank makes a larger profit than if we had $\gamma_2 < 0$. From an econometric viewpoint, it is often hard to give a conclusive statement about the stationarity property of a process. This difficulty leads us directly to the next question: How do the

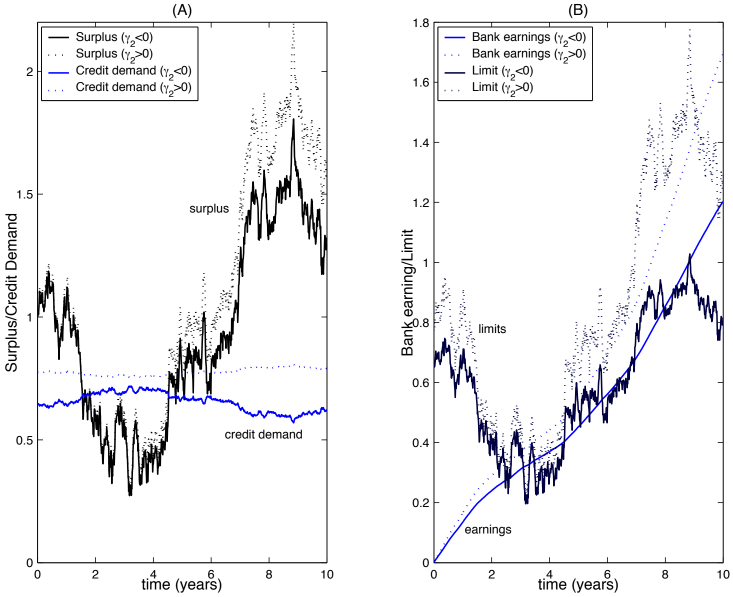


Figure 2: Optimal limit policy. We make the following assumptions: $\mu_0 = 2\%$, $\sigma_0 = 10\%$, $\sigma_1 = 30\%$, $\delta = 5\%$, $p = 1\%$, $\kappa = 0.8\%$. In order to generate $\gamma_2 < 0$, we set $\mu_1 = 5\%$. For $\gamma_2 > 0$, we simply set $\mu_1 = 10\%$. Panel (A) simulates one trajectory of the surplus given $\gamma_2 > 0$ (dotted line) and $\gamma_2 < 0$ (solid line) and its corresponding relative credit demand $c_t$. In Panel (B), we plot the absolute optimal limit policy corresponding to the surplus dynamics in Panel (A) for $\gamma_2 > 0$ (dotted line) and $\gamma_2 < 0$ (solid line), respectively. In Panel (B), we also plot the bank's (cumulative) profit resulting from the company's credit demand.

optimal policies change, when the bank has only partial information about the surplus and its dynamics?

# 3 Limit Policy with Partial Information

In practice, bank $B$ often has only partial information about the current surplus of company $C$.[3] Due to inaccuracies in the bank's measurement of the true surplus, the bank cannot measure $S_t$ itself, but a disturbed version of it:

$$\zeta_t = S_t + \text{“noise”}.$$

The bank has to make a credit limit assessment given a best estimate $\hat{S}$ of the company's surplus based on the observed signal $\zeta_t$. Such a best estimate is obtained by using a Kalman-Bucy filter (see Kalman (1960) and Kalman

[3] Such a model setup is relevant, e.g., if we interpret $c_t S_t$ as a deposit credit amount.

and Bucy (1961) for the original contributions). Intuitively, the bank filters away the noise from the system in an optimal way. To formalize the bank's behavior, we assume that the dynamics of the signal as observed by the bank follows the stochastic differential equation

$$(\mathbf{O}) \quad : \quad d\zeta_t = (A_0 + A_1 S_t)\, dt + B dZ_t.$$

The signal $\zeta_t$ can comprise such things as analysts' reports, share prices, press releases and so on. The state variable, i.e., the true surplus $S_t$ of company $C$, evolves according to

$$(\mathbf{U}) \quad : \quad dS_t = (\mu_0 + (\mu_1 - p)c_t^*)\, dt + (\sigma_0 + \sigma_1 c_t^*)\, dW_t,$$

with $c_t^*$ the company's optimal credit decision taking the bank's decision $\ell^*$ into account. We note that $\ell^*$ is now a function not only of $S_t$, but also of $\zeta_t$, the signal received by the bank. Then, company $C$ solves

$$(\mathbf{C2}): \quad J^C(S, t, \omega, \lambda) \ = \ \max_{c_t \in \hat{\mathcal{C}}} \mathbb{E}\left[\int_t^T e^{-\delta(s-t)} \left(\lambda S_s - \omega S_s^2\right) ds \,|\mathcal{F}_t\right]. \quad (5)$$

The above optimization problem is the same as the one in $(\mathbf{C1})$, with a subtle difference in the constraint set $\hat{\mathcal{C}}$: The upper bound for the current usage $c_t \leq \ell_t / S_t$ depends now also on the signal $\zeta_t$ through $\ell_t$. Using the same arguments as in Section 2, the optimal policy of company $C$ given the limit $\ell_t$ is

$$c_t^* = \ell_t / S_t. \quad (6)$$

To derive the bank's optimal limit policy, we first have to elaborate on the signal process and the best estimate for company's $C$ true surplus. To simplify our exposition, the two Brownian motions $W_t$ and $Z_t$ are assumed to be independent. We note that $\hat{S}_t$ is $\mathcal{G}_t$-measurable, where $\mathcal{G}_t$ is the $\sigma$-algebra generated by the Brownian motion $Z_t$. By saying that $\hat{S}_t$ is the "best guess" of the surplus we mean that

$$\int_\Omega |S_t - \hat{S}_t|^2 d\mathbb{P} = \mathbb{E}\left[|S_t - \hat{S}_t|^2\right] = \inf_Y \left\{\mathbb{E}\left[|\hat{S}_t - Y|^2\right] \big| Y \in L^2(\mathcal{G}, \mathbb{P})\right\}. \quad (7)$$

The equation for the unobservable surplus $(\mathbf{U})$ and the equation for the observable signal $(\mathbf{O})$ define a nonlinear optimal filtering problem. Deriving closed-form solutions for such filters is generally not possible. However, we can make use of the fact that the conditional distribution $F_{\zeta_t} = \mathbb{P}\left(S_t \leq k | \mathcal{G}_t\right)$ is $\mathbb{P}$-Gaussian.

**Proposition 4.** *Given equations* (**U**) *and* (**O**)*, the process* $\hat{S}_t$ *satisfying (7) is*

$$d\hat{S}_t = \left(\mu_0 + (\mu_1 - p)\hat{c}_t + \rho_t \frac{A_1^2}{B^2}(S_t - \hat{S}_t)\right) dt + \rho_t \frac{A_1}{B} dZ_t, \quad \hat{S}_0 = \mathbb{E}\left[S_0\right], \quad (8)$$

*with* $\rho_t$ *the solution to the Riccati equation*

$$\frac{d\rho_t}{dt} = (\sigma_0 + \sigma_1 \hat{c}_t)^2 - \rho_t^2 \frac{A_1^2}{B^2}, \quad \rho_0 = \mathbb{E}\left[(\hat{S}_0 - S_0)^2\right].$$

For a proof of the above proposition, we refer to Theorem 12.1 of Liptser and Shiryaev (2001). Equation (8) deserves some explanation. Recall that in Section 2 we define $c_t$ as a fraction of current surplus $S_t$. The product $c_t S_t$ makes up the amount of credit demand in absolute terms. Here, we claim that $S_t$ cannot be observed by the bank. However, what the bank observes is the actual credit demand $c_t S_t$. If the bank would already know the true value of $c_t$, then the bank would know the true value of $S_t$. To avoid such a trivial setting, we have to assume that not only $S_t$ is unobservable, but also $c_t$. Therefore, when the bank is faced with the credit demand $c_t S_t$, the bank first forms a best estimate $\hat{S}_t$ about the true surplus. From this estimate, the bank then infers the estimated value $\hat{c}_t$, since we require $c_t S_t = \hat{c}_t \hat{S}_t$. (See Figure 3.) It follows that in equation (8) the value of $\hat{c}_t$ enters the drift of $d\hat{S}_t$ and not the value of $c_t$. Furthermore, the amount $\hat{c}_t \hat{S}_t$ enters into the optimization problem (**B2**) below. Therefore, at first sight, it might look as if the bank is not concerned about the true value of $\hat{c}_t$ and $\hat{S}_t$. However, the dynamics of $\hat{S}_t$ enter the constraint set $\hat{\mathcal{B}}$. Hence, the optimization problem in (**B2**) is indeed different from the optimization problem (**B1**):

$$(\mathbf{B}2): \quad J^B(\hat{S}, t) \quad = \quad \max_{\ell_t \in \hat{\mathcal{B}}} \mathbb{E}\left[\int_t^T e^{-\delta(T-s)} \left(p\hat{c}_s \hat{S}_s - \kappa \ell_s\right) ds \,|\mathcal{G}_t\right], \quad (9)$$

where $\hat{\mathcal{B}}$ is same constraint set as for problem (**B**1), but with the surplus $S_t$ replaced by its best estimate $\hat{S}$ as given in equation (8).

The Hamilton-Jacobi-Bellmann (HJB) equation for (**B**2) is difficult to solve and possibly no closed-form solution can be obtained. One route to take would be to use numerical methods to obtain the optimal limit policy. However, at this stage we are not interested in quantitatively exact results, but want to learn more about qualitative features of the model. To this end, we use perturbation theory and expand around a point that has a concrete economic interpretation: Assume that bank $B$ is competent and the estimates of the surplus do not deviate drastically from the true surplus, at least in relative terms. Then $\left|\frac{S_t - \hat{S}_t}{S_t}\right|$ is small and close to 0. Moreover, as $\rho_t$ is highly non-

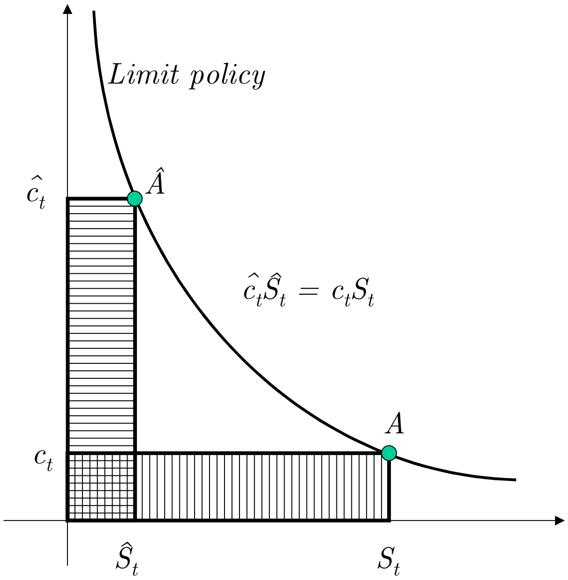


Figure 3: Credit demand: The bank observes the absolute value of the credit demand. This amount equals $c_tS_t = \hat{c}_t\hat{S}_t$. However, neither the true value of the surplus, $S_t$, nor the fraction $c_t$ are observable.

linear in $c_t$, we work with a first-order approximation in $c_t\sigma_1$. The results for the approximative strategies $c_t := c_t^{(1)*} + O\left(\left|\frac{S_t-\hat{S}_t}{S_t}\right|, (c_t\sigma_1)^2\right)$ and $\ell_t := \ell_t^{(1)*} + O\left(\left|\frac{S_t-\hat{S}_t}{S_t}\right|, (c_t\sigma_1)^2\right)$ are given in the next proposition.

**Proposition 5.** *Consider an economy with two players solving* (**B**2) *and* (**C**2)*, where player $B$ has only partial information on $C$'s surplus. With $B$ using $\zeta_t$ as a signal for the company's current surplus $S_t$, the strategies*

$$c_t^{(1)*} = \frac{\ell_t^{(1)*}}{S_t}, \ell_t^{(1)*} = \hat{\gamma}_1(t)\hat{S}_t + \hat{\gamma}_2(t)\hat{S}_t^2,$$

*are an asymptotic subgame perfect Nash equilibrium, where*

$$\hat{\gamma}_1(t) = \frac{B\sqrt{\delta}}{r_1(t)A_1}, \hat{\gamma}_2(t) = \frac{\mu_0\left(\mu_1 - p\right) - r_0(t)r_1(t)\frac{A_1^2}{B^2}\delta}{\sqrt{\delta}r_1(t)\frac{A_1}{B}\left(\mu_1 - p + \delta r_1(t)\frac{A_1^2}{B}\right)},$$

*and $r_0(t)$ and $r_1(t)$ are given in the Appendix, equation (A.7).*

From Proposition 5, we see that the differences between $\gamma_i$ and $\hat{\gamma}_i$ are given

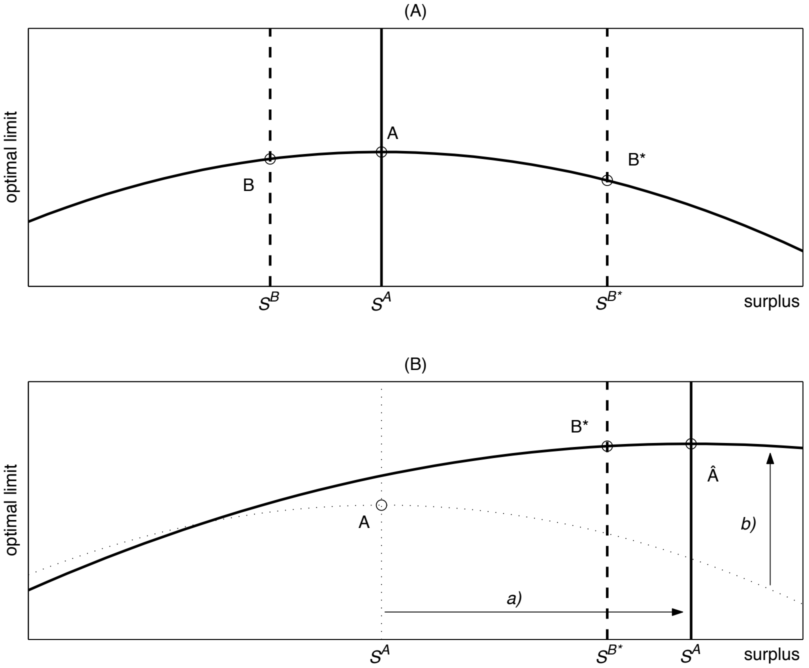


Figure 4: The influence of partial information. We assume: $\mu_0 = 2\%$, $\mu_1 = 5\%$, $\sigma_0 = 10\%$, $\sigma_1 = 30\%$, $\delta = 5\%$, $p = 1\%$. Panel (A) plots the complete information case. The concave curve describes the optimal limit as a function of the surplus. The straight line through point A represents the bank's satiation. The dotted lines represent the company's satiation, i.e., the line through point B when $\omega = 1.2$ and through point B$^*$ when $\omega = 1.125$. Panel (B) plots the case with partial information and for $\omega = 1.125$. For the signal process $\zeta_t$ we assume $A_1/B = 10$. The satiation for the bank is moved upwards to Â. In addition, the optimal limit policy is pushed upwards.

by changes in the volatility terms. More precisely, by substituting

$$\sigma_0 \to r_0(t)\frac{A_1}{B}, \quad \sigma_1 \to r_1(t)\frac{A_1}{B}, \tag{10}$$

in the expressions for $\gamma_i$ we obtain $\hat{\gamma}_i$. The value function of $J^B$ and $J^C$ are again quadratic, but in $\hat{S}_t$ and $S_t$ respectively.

Figure 4 clarifies the influence of first-order partial information in our model. In Figure 4 we plot the optimal limit policy as a function of the surplus $S$ when the mean $\mu_1$ is low such that the optimal limit policy in concave in $S$.

Panel (A) of Figure 4 plots the case when there is complete information. In addition to the optimal policy (the bold curve), we also plot the different satiation levels at which the company's and the bank's value functions are at a maximum. Whenever the optimal limit policy crosses a satiation level from the left, the company has no incentives to further increase her surplus, or, on the other hand, the bank has no incentives to further supply additional credit limits. Given the numerical parameter values, the satiation level for the bank in Panel (A) is the bold line that crosses the optimal limit policy in point A. For the company, we plot two different satiation levels. The dashed line that crosses the optimal policy at point B assumes $\omega = 1.2$, whereas the dashed line through point B$^*$ assumes $\omega = 1.125$, i.e., in the latter case, the company is slightly less risk-averse. We consider first the case with $\omega = 1.2$. The bank's satiation level is to the right of the company's satiation level. Therefore, the surplus realization will be point $S^B$, which corresponds to the company's satiation level. The bank will not be able to attain the surplus level that maximizes her value function. However, if the company has a risk aversion coefficient of $\omega = 1.125$, the bank reaches point $A$. The surplus will be at $S^A$ and, hence, maximizes the bank's value function. The company's satiation level, $S^{B*}$, will not be reached. Panel (B) of Figure 4 plots the optimal policy and the satiation levels in case of partial information and for $\omega = 1.125$. For comparison, we also plot the optimal policy and the bank's satiation level given complete information (dotted lines). When we introduce partial information, the volatilities $\sigma_0$ and $\sigma_1$ change according to equation (10). These changes induce two effects, indicated by the arrows in Panel (B) and labelled accordingly:

*a)* A right-shift of the bank's satiation level.

*b)* A shift of the optimal limit policy through a decrease in the curve's concavity.

Therefore, the change in the variances $\sigma_0$ and $\sigma_1$ moves the bank's satiation level from point A to Â. However, point Â is to the right of the company's satiation level (point B), that remains unaffected by the introduction of partial information. The resulting surplus, $S^{B*}$ satisfies now the company's but not the bank's satiation level. Therefore, the company is now better off than in the case with complete information. The bank offers more limit to the company and increases its potential exposure. Furthermore, the bank's regulatory costs (through $\kappa$) increase compared to the complete information regime. As a result, the company has a strong incentive to manipulate the signal dynamics $\zeta$. Indeed, by decreasing the ratio $A_1/B$, the company moves the bank's satiation level to the right. A decrease in $A_1/B$ occurs when the company increases the signal variance $B$ or decreases the $A_1$. The parameter $A_1$ determines the drift component in $\zeta_t$ that is proportional to the company's surplus. Therefore, a small $A_1$ makes the signal $\zeta_t$ less liable for predicting the true surplus $S_t$. We recall that the above results hold to up to first-order in $\left|\frac{S_t - \hat{S}_t}{S_t}\right|$, i.e., the true surplus is already near the bank's best guess. Therefore, considering higher-order terms, the distortion effects due to partial information would be even more accentuated.

# 4 Contracting under Partial and Incomplete Information

The above analysis of partial information shows a non-desirable feature from the bank's perspective. In this section, we discuss the situation in which the firm has to undertake some costly efforts, $\varepsilon$, to diminish the "noise" acting on the true state of her surplus. However, the bank cannot discriminate between firms which undertake high efforts to disclose their true surplus value from those that do not spend time and costs on this issue. Thus, in addition to having partial information on the state variable, the bank has incomplete information on whether the firm puts a high effort to diminish the signal's noise. Since a high effort is costly, the firm has no incentive to fully disclose her true surplus state. To remedy this situation, the bank has the possibility to set up a contract with the firm. This contract should be designed along the following lines:

i) The bank rewards the efforts to disclose the firm's surplus state by using a compensation scheme or contract. The effort function is assumed to be private information to the firm and, therefore, defines incomplete information for the bank.

ii) The contract should be incentive compatible for the firm and be at least as good as the next best opportunity.

The first requirement defines a contracting setup where the action of the firm is hidden to the bank (see Grossman and Hart (1983) for the general theory). Since the surplus cannot be observed by the bank, it is reasonable that also the firm's efforts to disclose the true surplus cannot be observed. This implies that a high value for $A_1/B$ can result either from low efforts or from high efforts, but the latter situation is much more probable. In the second requirement, we add incentive compatibility since we assume that information in the new contract variables is asymmetric between the firm and the bank. We assume quasi-linear utility functions for both the bank and the firm, which are thrice differentiable. For the bank, the formal model with incomplete and partial information reads:

$$(\widetilde{\mathbf{B}3}) \quad : \quad J^B(\hat{S},t) = \max_{\ell\in\hat{\mathcal{B}},K\in\mathcal{K}} \mathbb{E}^{\varepsilon}\left[\int_t^T e^{-\delta(T-s)}\left(p\hat{c}_s\hat{S}_s(\varepsilon) - \kappa\ell_s - K(\varepsilon)\right) ds|\mathcal{G}_t\right]. \tag{11}$$

subject to

$$\begin{aligned}
\varepsilon^*, c_t^* &\in \operatorname{argmax}_{c_t \in \hat{\mathcal{C}}, \varepsilon} \mathbb{E}^\varepsilon \left[ \int_t^T e^{-\delta(T-s)} \left( \lambda S_s(\varepsilon) - \omega S_s^2(\varepsilon) + K(\varepsilon) \right) ds \,|\mathcal{F}_t \right], && (12) \\
\bar{u} &\leq \mathbb{E}^\varepsilon \left[ \int_t^T e^{-\delta(T-s)} \left( \lambda S_s(\varepsilon) - \omega S_s^2(\varepsilon) + K(\varepsilon) \right) ds \,|\mathcal{F}_t \right], && (13) \\
dS_t &= (\mu_0 + (\mu_1 - (p + \varepsilon))c_t)dt + (\sigma_0 + \sigma_1 c_t) dW_t. && (14)
\end{aligned}$$

Choosing an effort to disclose the surplus state is costly for the firm, but, on the other side, also beneficial. This is reflected by the dynamics of the modified non-observable surplus in condition (14) that contains both the reward $K(\varepsilon)$ through $c_t$ and a cost function for the chosen effort. For simplicity, the cost function is assumed to be the identity function. The reward adjusts the return $\mu_1$ and the cost function adjusts the price $p$ paid for the credit usage. The incentive constraint (12) ensures that the firm always chooses an effort which is in her self-interest. The constraint (13) is the participation or individual-rationality constraint of the firm. It states that the firm accepting a contract is not better off choosing any alternative contract that gives the firm a constant utility $\bar{u}$.

To formally define such an implementable allocation, we assume that the effort level $\varepsilon$ ranges in a closed and compact interval $[\underline{\varepsilon}, \overline{\varepsilon}]$. The bank has the prior cumulative distribution function $F^\varepsilon$ with differentiable density $f^\varepsilon$, such that the density is strictly positive for $\varepsilon$. Therefore, the expectation in the bank's objective function given in $(\widetilde{\mathbf{B}3})$ taken under the product measure of the distribution of firm types $\varepsilon$ and the surplus distribution.

The program $(\widetilde{\mathbf{B}3})$ generalizes the standard theory of contracts with hidden information[4] in two respects: First, $(\widetilde{\mathbf{B}3})$ is a dynamic program. Second, the state variable $S$ is not observable for the bank. Instead, there is a noisy signal, from which a best guess on the true state can be extracted. This defines partial information about the state variable.

We next provide a precise model formulation and the corresponding solution concept step-by-step. This is done in *Step I* to *Step VI*.

*Step I*: The bank estimates the surplus state $\hat{S}$ using the non-linear filter technique presented in Section 3. Repeating the calculations of Section 3, the dynamics of the bank's best guess for the firm's surplus is obtained as

$$d\hat{S}_t = \left( \mu_0 + (\mu_1 - (\varepsilon + p))\hat{c}_t + \rho_t \frac{A_1^2}{B^2}(S_t - \hat{S}_t) \right) dt + \rho_t \frac{A_1}{B} dZ_t, \qquad (15)$$

where the solution for $\rho_t$ is given in equation (A.7).

[4] See Mirrless (1971), Grossman and Hart (1983), Fudenberg and Tirole (1993).

*Step II*: We restrict the agents' optimal decisions by imposing two conditions, $\hat{c}_t = \frac{\hat{\ell}_t}{\hat{S}_t}$ and $\hat{c}_t \hat{S}_t = c_t S_t$. These two conditions allow us to eliminate the decision variable $\ell_t$. The first condition is based on the observation that, under incomplete information, it should never pay to offer a higher limit amount than the one calculated by using the best guess on the firm's surplus. The rationale for the second condition is that, given a firm demands credit, the bank cannot decide on $c_t$ and $S_t$ separately due to partial information. The bank only observes $c_t S_t$.

*Step III*: The contract allocation consisting of the chosen credit demand $c(\varepsilon)$ and the payment $K(\varepsilon)$ have to satisfy the following requirements: The payments resulting from the contract have to be feasible and satisfy the individual rationality constraint. Given the set of feasible allocations, the bank finally chooses the contract payments in the class with the highest expected payoff.

**Definition 1.** *A decision function $c(\varepsilon)$ is implementable if there exists a payment $K(\varepsilon)$ such that the type-contingent allocation $y(\varepsilon) = (c(\varepsilon), K(\varepsilon))$ satisfies the constraint*

$$u(y(\varepsilon), S(\varepsilon)) \geq u(y(\hat{\varepsilon}), S(\varepsilon)) \ , \ \forall (\varepsilon, \hat{\varepsilon}) \in [\underline{\varepsilon}, \overline{\varepsilon}] \times [\underline{\varepsilon}, \overline{\varepsilon}], \tag{16}$$

*where the function $u(\cdot)$ is given by:*

$$u(y(\varepsilon), S(\varepsilon)) \ = \ \int_t^T e^{-\delta(T-s)} \left( \lambda S_s(\varepsilon) - \omega S_s^2(\varepsilon) + K(\varepsilon) \right) ds. \tag{17}$$

The constraint in (16) is the standard form for the constraint given in equation (12) of problem ($\widetilde{\mathbf{B}3}$). Basically, it states that the firm optimally consumes the credit corresponding to her effort revealed to the bank. Whenever the revealed effort $\hat{\varepsilon}$ and the true effort $\varepsilon$ disagree, the utility for the firm is reduced. Equivalently to equation (17), we will define the bank's utility function as

$$v(y(\varepsilon), S(\varepsilon)) = \int_t^T e^{-\delta(T-s)} \Big( (p - \kappa) \hat{c}_s(\varepsilon) \hat{S}_s(\varepsilon) - K(\varepsilon) \Big) \ ds. \tag{18}$$

Both functions $u(\cdot), v(\cdot)$, and the decision variable $\hat{c}(\varepsilon)$ are assumed to be sufficiently smooth.

The condition (16) is difficult to handle in dynamic models with a continuous effort space. Therefore, at the price of some extra a-priori assumptions on the functional forms, we introduce some additional necessary conditions.

**Proposition 6.** *Consider the problem* $(\widetilde{\mathbf{B}3})$ *and (17). The condition*

$$\frac{\partial \hat{c}_t(\varepsilon)}{\partial \varepsilon}\frac{\partial^2 u(y(\varepsilon),S(\varepsilon))}{\partial\varepsilon\partial c_t} \geq 0 \tag{19}$$

*is sufficient for a piecewise function* $\hat{c}_t(\varepsilon)$ *to be implementable.*

Due to the partial information structure, we have two state variables $S$ and $\hat{S}$. The usual approach in contract theory to deal with the hidden action $\varepsilon$ is not applicable here, since the asymmetric information is related to only one state variable. We therefore assume that the bank is choosing an $\hat{S}$-optimal policy. The definition of this policy is given below.

**Definition 2.** *An* $\hat{S}$*-optimal policy for the bank consists of an implementable allocation* $y = (c(\varepsilon), K(\varepsilon))$ *and satisfies the individual rationality constraint (13) in problem* $(\widetilde{\mathbf{B}3})$ *where the bank always assumes that* $\hat{S}$ *is the true state.*

*Step IV*: With the above prerequisites, the $\hat{S}$-optimal program can be formulated as

$$(\mathbf{B3}) \quad : \quad J^B(\hat{S},t) = \max_{\hat{c}(\varepsilon),K(\varepsilon),\varepsilon} \mathbb{E}^\varepsilon\left[v(y(\varepsilon),\hat{S}(\varepsilon))\big|\mathcal{G}_t\right]. \tag{20}$$

subject to (15) and

$$\begin{aligned}\mathbb{E}^\varepsilon\left[u(y(\varepsilon),\hat{S}(\varepsilon))\,|\mathcal{F}_t\right] &\geq \mathbb{E}^\varepsilon\left[u(y(\hat{\varepsilon}),\hat{S}(\varepsilon))\,|\mathcal{F}_t\right]\;,\;\forall(\varepsilon,\hat{\varepsilon})\in[\underline{\varepsilon},\overline{\varepsilon}]\times[\underline{\varepsilon},\overline{\varepsilon}] & (21)\\ \bar{u} &\leq \mathbb{E}^\varepsilon\left[u(y(\varepsilon),\hat{S}(\varepsilon)\,|\mathcal{F}_t\right], & (22)\end{aligned}$$

where the solution for $\rho_t$ is given in equation (A.7) and $\hat{S}_0 = \mathbb{E}^\varepsilon[S_0]$.

*Step V*: The next step to obtain a solution for (20) is to eliminate the transfer payment $K(\varepsilon)$ in the optimization program.

**Proposition 7.** *Subject to (15), the optimization program (20) is equivalent to*

$$J^B(\hat{S},t) \;=\; \max_{\hat{c}(\varepsilon)}\mathbb{E}^\varepsilon\left[\int_{\underline{\varepsilon}}^{\overline{\varepsilon}}\left(v_1(c(\varepsilon),\hat{S}(\varepsilon)) + u_1(\hat{c}(\varepsilon),\hat{S}(\varepsilon)) - \frac{1-F^\varepsilon}{f^\varepsilon}\frac{\partial u_1(\hat{c}(\varepsilon),\hat{S}(\varepsilon))}{\partial\varepsilon}\right)f^\varepsilon d\varepsilon\big|\mathcal{G}_t\right]$$

*where* $u_1(\hat{c}(\varepsilon),\hat{S}(\varepsilon)) = u(y(\varepsilon),\hat{S}(\varepsilon)) - \int_t^T e^{-\delta(T-s)}K(\varepsilon)ds$ *and* $v_1(\hat{c}(\varepsilon),\hat{S}(\varepsilon)) = v(y(\varepsilon),\hat{S}(\varepsilon)) + \int_t^T e^{-\delta(T-s)}K(\varepsilon)ds$*. The contract is given by*

$$K^*(\varepsilon) = e^{\delta(T-t)}\frac{\partial}{\partial t}\left(u_1(\hat{c}(\varepsilon),S(\varepsilon)) - \int_{\underline{\varepsilon}}^{\varepsilon}\frac{\partial u_1\left(\hat{c}(\tilde{\varepsilon}),S(\tilde{\varepsilon})\right)}{\partial\tilde{\varepsilon}}d\tilde{\varepsilon}\right).$$

|  |  | time |  |
|---|---|---|---|
|  |  | static | dynamic |
| information | complete/partial | $(A)$ | $(A)+(B_1)$ |
|  | incomplete/partial | $(A)+(C)$ | $(A)+(B_1)+(C)+(B_2)$ |

Table 1: Static and dynamic optimization, and different information structures.

Hence, if we know the optimal policy $c^*$, we can explicitly determine the optimal contract $K^*(\varepsilon)$.

*Step VI*: The final step is to solve for $c^*$. We content ourselves with stating the optimality condition,

$$\underbrace{\mathcal{A}_{\hat{S}} J^B(\hat{S},t)}_{(B_1)} + \underbrace{\frac{\partial J^B(\hat{S},t)}{\partial \hat{c}} + \frac{\partial J^C(\hat{S},t)}{\partial \hat{c}}}_{(A)} = \underbrace{\frac{1-F(\varepsilon)}{f(\varepsilon)} \frac{\partial^2 u_1(\hat{c}(\varepsilon),\hat{S}(\varepsilon))}{\partial\varepsilon\partial\hat{c}}}_{(C)} + \underbrace{\Delta\mathcal{A}_{\hat{S}} J^B(\hat{S},t)}_{(B_2)},$$

and categorize these components in Table 1. By inspection, the bank faces a tradeoff between maximizing the joint surplus, given by the expression $(A)$, and appropriating the firm's information rent, represented by $(C)$. The term $(C)$ is solely determined by factors depending on the characteristics of the firm. If we were in a static economy, the optimum would be obtained when an increase in the joint surplus equals the expected increase in the company's rent.

For $\varepsilon = \overline{\varepsilon}$, the term $(C)$ is just zero and the extraction of the company's effort is not a concern, i.e., only $(A)$ is maximized. Because $\frac{1-F(\varepsilon)}{f(\varepsilon)} \geq 0$, for all other levels of effort $\varepsilon$, the impact of $(C)$ on the bank's marginal utility function depends on the sign of the cross-derivative $\frac{\partial^2 u_1(\hat{c}(\varepsilon),\hat{S}(\varepsilon))}{\partial\varepsilon\partial\hat{c}}$. Since it would be against economic intuition if $\frac{\partial\hat{c}(\varepsilon)}{\partial\varepsilon} < 0$, we have from equation (19) that the sign of $\frac{\partial^2 u_1(\hat{c}(\varepsilon),\hat{S}(\varepsilon))}{\partial\varepsilon\partial\hat{c}} \geq 0$ is positive, i.e., a higher surplus makes a higher demand for credit more desirable. Therefore, the term $(C)$ is strictly positive for $\varepsilon < \overline{\varepsilon}$.

The effect of the effort on the magnitude of $(C)$ is less clear. As noted in Fudenberg and Tirole (1993) (chapter 7), the optimal decision obtained by ignoring the monotonicity constraint satisfies monotonicity, if we replace the monotonicity condition $\frac{\partial\hat{c}(\varepsilon)}{\partial\varepsilon} > 0$ with the assumption $\frac{\partial^3 u_1(\hat{c}(\varepsilon),\hat{S}(\varepsilon))}{\partial\varepsilon^2\partial\hat{c}} \leq 0$. Therefore, with the sufficient condition $\frac{\partial^3 u_1(\hat{c}(\varepsilon),\hat{S}(\varepsilon))}{\partial\varepsilon^2\partial\hat{c}} \leq 0$, the term $(C)$ induces the bank to attract debtors with a low effort $\varepsilon$, if and only if the hazard rate $\frac{f(\varepsilon)}{1-F(\varepsilon)}$ satisfies $\frac{d}{d\varepsilon}\left(\frac{f(\varepsilon)}{1-F(\varepsilon)}\right) \geq 0$. Bagnoli and Bergstrom (1989) show that the latter condition is equivalent to $f(\varepsilon)$ being log-concave on $[\underline{\varepsilon}, \overline{\varepsilon}]$. Log-concavity is fulfilled for distributions such as the uniform, normal, logistic, and $\chi^2$-distribution, but in general not, e.g., for the Student's-t distribution. Therefore, monotonicity of the cross-derivative and a log-concave distribution function for $\varepsilon$ on $[\underline{\varepsilon}, \overline{\varepsilon}]$ make low efforts more desirable for the bank in order to increase the impact of $(C)$ on her value function. In contrast, a non-monotone hazard rate will prevent the bank to attract firms with low effort. We conclude that, from the bank's perspective, the optimal effort level to reveal the true surplus strongly depends on the properties of the bank's prior distribution of the firm's type (effort). The bank tries to obtain a lower effort level, if the prior distribution is log-concave, and a higher effort level if the hazard rate is decreasing in $[\underline{\varepsilon}, \overline{\varepsilon}]$. The optimal effort level not only depends on the behavior of the quotient $\frac{1-F(\varepsilon)}{f(\varepsilon)}$, but also on its influence on other terms. When we move from a static to a dynamic setup, there are two intertemporal components entering the scene, $(B_1)$ and $(B_2)$. $(B_1)$ is invariant to the possibility of incomplete information. Contrary to $(B_2)$, it does not depend on the effort $\varepsilon$. The term $(B_2)$ serves as a correction term for the standard generator $\mathcal{A}_{\hat{S}}$. The generator decreases when the effort level decreases, which reduces the bank's value function. Therefore, $(B_2)$ can be interpreted as a dynamic hedging component against incomplete information on the firm's effort $\varepsilon$, and $(B_1)$ as the dynamic hedging component against the noisy signal.

# 5 Implication for Credit Risk Management

## 5.1 Model with Full Information

From the optimal policies derived in Proposition 2, the marginal utility of an additional dollar surplus is always larger for the firm than for the bank. The discount factor and the additional risk of a new investment uniquely determine the difference between the two marginal utilities. When the firm and the bank become more impatient, the increase in the marginal utility from an additional investment turns out to be larger for the bank than for the firm. Similarly, if the risk of new investments decreases to zero, the utility from an additional dollar surplus increases unlimitedly for the bank, whereas for the firm the additional utility tends to zero. Therefore, given full information, banks should finance long-term investments with low risks.

To what extend should this long-term financing grow? The sign of $\gamma_2$ gives an unambiguous answer. Given that $\gamma_2 > 0$, the optimal limit policy is not bounded from above. In the other case, once the surplus has reached a critical level, the bank should stop to assign additional credit limits. Under what conditions does the sign of $\gamma_2$ turn positive or negative? Separating the surplus dynamics $S$ into a sum $S = S_0 + S_1$, where $S_0$ denotes the surplus when $c^* = 0$ and $S_1 = S - S_0$, then the inequality $\gamma_2 > 0$ holds if

$$\mathbb{E}\left(dS_0\right)\mathbb{E}\left(dS_1/c^*\right) > \delta\sqrt{\mathrm{Var}\left(dS_0\right)}\sqrt{\mathrm{Var}\left(dS_1/c^*\right)}, \tag{23}$$

given that the firm's additional return on investment is larger than the bank's cost of capital, i.e., $\mu_1 > p$.

The inequality (23) allows us to derive several implications. First, the lower the risk of new investments, the more likely inequality (23) holds. As a consequence, the bank should finance any credit demands. Second, if the bank and the firm are increasingly impatient, then an optimizing bank will bound the total limit usage once the surplus of the firm reached a given known level.

So far, no distinction was made whether a firm asked for one million or one billion dollars. The price of one dollar credit $p$ was assumed to be constant. If we consider a small firm asking for a one million dollar credit and a large firm asking for a one billion dollar credit, then a discount, i.e., $p_1 > p_2$, is a reasonable assumption. Since $\gamma_2$ is strictly decreasing in $p$, the inequality (23) holds more likely for the small than for the large firm. Hence, if we charge a higher price to the small firm, it is rational to bound the total credit limit of the large firm for a known surplus level. Put differently, a large firm cannot expect both a discount and unlimited resources from the bank's side.

## 5.2 Model with Partial Information

In this model, the bank has not full information about the firms surplus. Instead, the bank observes a noisy signal. The firm can exploit the bank's lack of full information through manipulating the signal and thereby rising the credit limits. Basically, this can be done in two ways. First, the firm can produce a highly noisy signal, since it is in the firm's self-interest if a large amount of noise hides their true firm value or surplus state. Second, the firm may behave in such a way so as to weaken the correlation between the nonobservable surplus and the observable signal. To the extreme, the firm can even try to make the correlation negative.

However, it is not optimal for all firms to behave in this way. Consider a firm whose risk and return properties satisfies the analogue of inequality (23) for the partial information case. Then, the firm has a less pronounced incentive to produce information distortions, since in any case it is optimal for the bank to cover any credit demand. Therefore, only a firm with less favorable risk and return properties will have an incentive to manipulate the signal as described above.

## 5.3 Model with Partial and Incomplete Information

If the bank faces partial information, the bank will counteract the firm's incentives to produce noisy signals. Since it is difficult to assess the firm's effort to disclose her surplus status, we assume that the bank has only incomplete information about the firm's effort. Therefore, the bank has to work out a clever contracting scheme such that it is optimal for the firm to reveal the true effort. Such a procedure leads to second-best contracts, i.e., the resolution of the incompleteness leads to an optimal decision that is costly and less efficient than the policy with complete information. Therefore, given the possibility of efficiency losses, institutions should decide whether it is worth to work out an incentive compatible contract to meet the firm's credit demand. The alternative would be to simply refuse the credit demand.

If a bank decides to enter into the contract, a key model risk is the prior distribution function used to assess the 'type' of the counterparty. This probability function measures the probability that the firm chooses a low, a medium, or a high effort level. The assessment of these probabilities has far reaching consequences, since it basically determines which types of firms the bank attracts. This attraction will finally determine the credit portfolio's composition.

As an example, we consider a bank that assumes the distribution to be uniform for one client and Student's $t$ distributed for a second one. To simplify the analysis, we consider a static setup. At the optimum, up to constants, the bank's value function $J^B$ and firm's value function $J^C$ satisfy

$$J^B = -J^C + (\bar{\epsilon} - \epsilon)\,\frac{\partial^2 u_1(\hat{c}(\varepsilon), \hat{S}(\varepsilon))}{\partial\varepsilon\partial\hat{c}}, \tag{24}$$

for the uniform distribution with $\bar{\epsilon}$ the maximum effort level, and

$$J^B = -J^C + \frac{(1+\epsilon^2)\,\left(\pi - 2\ \arccos(\sqrt{\frac{1}{1+\bar{\epsilon}^2}})\mathrm{sgn}(\bar{\epsilon})\right)}{2}\,\frac{\partial^2 u_1(\hat{c}(\varepsilon), \hat{S}(\varepsilon))}{\partial\varepsilon\partial\hat{c}}\ , \tag{25}$$

for the Student's $t$ distribution with one degree of freedom. For the uniform distribution, since equation (25) is monotone increasing in $\epsilon$, the bank prefers firms with low effort level. For the Student's $t$ distribution, the bank prefers firms with a high effort level. Hence, if the firm has no information at hand (uniform case), different types of counterparties enter the bank's credit portfolio. In the worst case, wrong prior beliefs will lead to an unintended composition of the bank's credit portfolio.

# 6 Conclusion

The increasing importance of limit management is related to the deterioration of credit markets during recent years, which makes short-term decisions an indispensable tool for adequate risk management. We analyze limit management under different information setups. We start with a model, where both the bank and the firm have complete information about the firm's state variable. Since small variations in the model parameters lead to large differences in the optimal limit policy and since banks often has to assess a credit limit with partial information on the firm's true surplus we are led to consider a partial information model using noisy signals. But a noisy signal for the firm's true surplus introduces some undesirable effects for the bank since incentives of the firm are distorted. However, these incentive distortions can be reduced by implementing an incentive optimal contract for the firm to put some effort into disclosing her true surplus. This, in turn, gives rise to an optimization problem in the presence of incomplete information. Setting up an incentive-compatible contract may turn out to be costly. Furthermore, the success of such contracting strongly depends on the correct assessment of the prior distribution of counterparty qualities.

# References

Bagnoli, M. and Bergstrom, T.: 1989, Log-concave probability and its applications, *Working paper.*

Black, F. and Cox, J.: 1976, Valuing corporate securities: Some effects of bond indenture provision, *Journal of Finance* **31**(2), 351–367.

Das, S. (ed.): 2000, *Credit Derivatives and Credit Linked Notes: Trading and Management of Credit and Default Risk*, John Wiley and Sons.

Egloff, D., Leippold, M. and Vanini, P.: 2003, Integrated credit and business risk portfolio management, *Technical report*, University of Southern Switzerland.

Francis, J., Frost, J. and Whittaker, J.: 1999, *Handbook of Credit Derivatives*, Irwin/McGraw-Hll, New York.

Fudenberg, D. and Tirole, J.: 1993, *Game Theory*, The MIT Press, Boston, Chicago.

Giesecke, K. and Weber, S.: 2002, Credit contagion and aggregate losses, *Working paper*, Cornell University and Technische Universität Berlin.

Grossman, S. and Hart, O.: 1983, An analysis of the principal-agent problem, *Econometrica* **51**(1), 7–45.

Holmström, B. and Milgrom, P.: 1987, Aggregation and linearity in the provision of intertemporal incentives, *Econometrica* **55**, 303–328.

Isaacs, R.: 1954, Differential games, i, ii, iii, iv, *Reports rm-1931, 1399, 1411, 1486*, Rand Corporation.

Kalman, R.: 1960, A new approach to linear filtering and prediction problems, *Journal of Basic Engineering* **82**, 35–45.

Kalman, R. and Bucy, R.: 1961, New results in linear filtering and prediction theory, *Journal of Basic Engineering* **83**, 95–108.

Leippold, M., Trojani, F. and Vanini, P.: 2004, A geometric approach to multiperiod mean variance optimization of assets and liabilities, *Journal of Economic Dynamics and Control* **28**(6), 1079–1113.

Li, D.: 2000, On default correlation: A copula approach, *Journal of Fixed Income* **9**, 43–54.

Lintner, J.: 1956, Distribution of incomes of corporations among dividends, retained earnings, and taxes, *American Economic Review* **46**, 7113.

Liptser and Shiryaev, A.: 2001, *Statistics of Random Processes*, Applications of Mathematics, second edn, Springer-Verlag, Berlin, Heidelberg, New York.

Lucas, D.: 1995, Default correlation and credit analysis, *Journal of Fixed Income* **4**, 76–87.

Markowitz, H.: 1952, Portfolio selection, *Journal of Finance* **7**, 77–91.

Merton, R.: 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* **2**(2), 449–470.

Mirrless, J.: 1971, An exploration in the theory of optimum income, *Review of Economic Studies* **38**, 175–208.

Nelken, I.: 1999, *Implementing Credit Derivatives. Strategies and Techniques for Using Credit Derivatives in Risk Management*, Irwin/McGraw-Hill, New York.

Zhou, X. and Li, D.: 2000, Continuous-time mean-variance portfolio selection: A stochastic LQ framework, *Applied Mathematics and Optimization* **42**(1), 19–33.

# Appendix

*Proof of Proposition 2 and Proposition 3.* For company $C$, the value function $J^C$ must satisfy $0 = \max_{c_t \in \mathcal{C}} \left( e^{-\delta(T-t)} \left( \lambda S_t - \omega S_t^2 \right) + \mathcal{L} J^C - \phi \left( c_t S_t - \ell_t \right) \right)$, where $\phi$ is the Lagrange multiplier and $\mathcal{L}$ is the extended generator of $S$. Solving for $c^*$ and making the Ansatz $J^C(S,t) = e^{-\delta(T-t)} \left( k_0 + k_1 S_t + \frac{1}{2} k_2 S_t^2 \right)$, we obtain the parameters $k_i$ as

$$
\begin{aligned}
k_0 &= \frac{k_1^2 \left(\mu_1 - p\right)^2 - 2k_2 \phi \sigma_2^2 \ell_t + 2k_1 k_2 \sigma_1 \left( \left(\mu_1 - p\right) \sigma_0 - \mu_0 \sigma_1 \right)}{2k_2 \delta \sigma_1^2}, \\
k_1 &= k_2 \sigma_1 \frac{\left(\phi - k_2 \left(\mu_1 - p\right)\right) \sigma_0 + \left(\lambda + k_2 \mu_0\right) \sigma_1}{\left(\phi - k_2 \left(\mu_1 - p\right)\right) \left(p - \mu_1\right) - k_2 \delta \sigma_1^2}, \\
k_2 &= \frac{\phi(\mu_1 - p) - \sigma_1^2 \omega + \sigma_1 \sqrt{\delta \phi^2 - 2\phi\omega(\mu_1 - p) + \sigma_1^2 \omega^2}}{(\mu_1 - p)^2 - \delta \sigma_1^2}.
\end{aligned}
$$

Determining the Lagrange multiplier and substituting into the expression for $c^*$, we get $c_t^* = \frac{\ell_t}{S_t}$. For the bank, the HJB equation is $0 = \max_{\ell \in \mathcal{B}} \left( e^{-\delta(T-t)} \left( p c_t S_t - \kappa \ell_t \right) + \mathcal{L} J \right)$. With $J^B(S,t) = e^{-\delta(T-t)} (p - \kappa) \left( b_0 + b_1 S_t + \frac{1}{2} b_2 S_t^2 \right)$, the coefficients $b_i$ are obtained as

$$
\begin{aligned}
b_0 &= \frac{\left(p - \mu_1\right) \left(\delta \sigma_0^2 - \mu_0^2\right) \left( \mu_1 - p + 2\sqrt{\delta} \sigma_1 \right) - 2\delta \mu_0 \sigma_1^2 \left( \mu_0 + \sqrt{\delta} \sigma_0 \right)}{2\delta^2 \left( \mu_1 - p + \sqrt{\delta} \sigma_1 \right)^3}, && \text{(A.1)} \\
b_1 &= \sigma_1 \frac{\mu_0 + \sqrt{\delta} \sigma_0}{\sqrt{\delta} \left( \mu_1 - p + \sqrt{\delta} \sigma_1 \right)^2}, && \text{(A.2)} \\
b_2 &= -\frac{1}{2 \left( \mu_1 - p + \sqrt{\delta} \sigma_1 \right)}. && \text{(A.3)}
\end{aligned}
$$

Plugging these results into the expression for $\ell_t^*$, we obtain $\gamma_1$ and $\gamma_2$ as claimed in the proposition. Finally, plugging the optimal limit policy into the value function of $C$, we obtain the parameters $k_i$ as

$$
\begin{aligned}
k_0 &= -\frac{2k_1 \left( \gamma_1 (\mu_1 - p) + \mu_0 \right) + k_2 \left( \sigma_0 + \gamma_1 \sigma_1 \right)^2}{2\delta}, && \text{(A.4)} \\
k_1 &= -\frac{\lambda + k_2 \left( \mu_0 + \gamma_2 \sigma_0 \sigma_1 + \gamma_1 \left( \mu_1 - p + \gamma_2 \sigma_1^2 \right) \right)}{\delta + \gamma_2 (\mu_1 - p)}, && \text{(A.5)} \\
k_2 &= \frac{2\omega}{\delta + \gamma_2 \left( 2(\mu_1 - p) + \gamma_2 \sigma_1^2 \right)}. && \text{(A.6)}
\end{aligned}
$$

This concludes the proof. □

*Proof of Proposition 5.* The HJB equation of the company has solution $J^C(S,t) = e^{-\delta(T-t)} \left( \hat{k}_0 + \hat{k}_1 S_t + \frac{1}{2} \hat{k}_2 S_t^2 \right)$. Evaluating the Lagrange multiplier and plugging it into the optimal policy, yields $c^* = \ell_t / S_t$. The bank's HJB equation is given by

$$0 = \max_{\ell \in \mathcal{B}(\hat{S})} \left( e^{-\delta(T-t)} \left( p\hat{c}_t \hat{S}_t - \kappa \ell_t \right) + \hat{\mathcal{L}} J^B \right),$$

where $\hat{\mathcal{L}}$ is the extended generator of the filtered process given in (8). From Proposition 4, $\rho_t$ is obtained as

$$\rho_t = \frac{\left(\sigma_0 + \sigma_1 \hat{c}_t\right) B^2}{A_1^2} \times \frac{\frac{A_1^2}{B^2}\rho_0 - \left(\sigma_0 + \sigma_1 \hat{c}_t\right) + e^{2\left(\sigma_0 + \sigma_1 \hat{c}_t\right)\frac{A_1^2}{B^2}t}\left(\frac{A_1^2}{B^2}\rho_0 + \left(\sigma_0 + \sigma_1 \hat{c}_t\right)\right)}{\left(\sigma_0 + \sigma_1 \hat{c}_t\right) - \frac{A_1^2}{B^2}\rho_0 + e^{2\left(\sigma_0 + \sigma_1 \hat{c}_t\right)\frac{A_1^2}{B^2}t}\left(\frac{A_1^2}{B^2}\rho_0 + \left(\sigma_0 + \sigma_1 \hat{c}_t\right)\right)}.$$

The above equation is non-linear in $\hat{c}_t$ and has no closed-form solution. The parameter $\hat{c}_t$ always appears in $\rho_t$ scaled by $\sigma_1$. We argue that $\hat{c}_t \sigma_1$ is usually small and linearize $\rho_t$ by using a first-order approximation around $\hat{c}_t \sigma_1 = 0$. Then, $\rho_t^{(1)}$ defined by $\rho_t = \rho_t^{(1)} + O\left( (\hat{c}_t \sigma_1)^2 \right)$ is

$$\begin{aligned}
\rho_t^{(1)} \quad &= \quad \frac{\sigma_0 B^2}{A_1^2} \times \frac{\frac{A_1^2}{B^2}\rho_0 \cosh\left[\frac{A_1^2}{B^2}\sigma_0 t\right] + \sigma_0 \sinh\left[\frac{A_1^2}{B^2}\sigma_0 t\right]}{\sigma_0 \cosh\left[\frac{A_1^2}{B^2}\sigma_0 t\right] + \frac{A_1^2}{B^2}\rho_0 \sinh\left[\frac{A_1^2}{B^2}\sigma_0 t\right]} + \frac{\sigma_1 B^2}{A_1^2} \\
&\times \frac{e^{4\frac{A_1^2}{B^2}\sigma_0 t}\left(\frac{A_1^2}{B^2}\rho_0 + \sigma_0\right)^2 - \left(\frac{A_1^2}{B^2}\rho_0 - \sigma_0 \frac{A_1^2}{B^2}\right)^2 - 4\frac{A_1^2}{B^2}\sigma_0 e^{2\frac{A_1^2}{B^2}\sigma_0 t}\left(\rho_0 + \frac{A_1^4}{B^4}\rho_0^2 t - \sigma_0^2 t\right)}{\left(\sigma_0 - \frac{A_1^2}{B^2}\rho_0 + e^{2\frac{A_1^2}{B^2}\sigma_0 t}\left(\frac{A_1^2}{B^2}\rho_0 + \sigma_0\right)\right)^2}\hat{c}_t, \\
&=: \quad r_0(t) + r_1(t)\hat{c}_t. \qquad\qquad (A.7)
\end{aligned}$$

Rewrite the HJB equation as

$$\begin{aligned}
0 \quad = \quad & \left( \mu_0 + r_0(t)\frac{A_1^2}{B^2}\left(S_t - \hat{S}_t\right) + \left(\mu_1 - p + r_1(t)\frac{A_1^2}{B^2}\left(S_t - \hat{S}_t\right)\right)\hat{c}_t \right) J^B_{\hat{S}} \\
& + \frac{A_1^2}{2B^2}\left(r_0(t) + r_1(t)\hat{c}_t\right)^2 J^B_{\hat{S}\hat{S}} + p\hat{c}_t\hat{S}_t - \kappa\ell_t + J^B_t, \qquad (A.8)
\end{aligned}$$

which is hard to solve. We therefore follow a perturbative approach and assume that $\left|\frac{S_t - \hat{S}_t}{\hat{S}_t}\right|$ is small. Expanding in $\left|\frac{S_t - \hat{S}_t}{\hat{S}_t}\right|$ around zero, we obtain

$$0 \quad = \quad p\hat{c}_t\hat{S}_t - \kappa\ell_t + J^B_t + \left(\mu_0 + (\mu_1 - p)\hat{c}_t\right) J^B_{\hat{S}} + \frac{A_1^2}{2B^2}\left(r_0(t) + r_1(t)\hat{c}_t\right)^2 J^B_{\hat{S}\hat{S}}.$$

Now, determining the $J^B$ function is equivalent to the proof of Proposition 2, but with $\sigma_0$ and $\sigma_1$ replaced by $r_0(t)\frac{A_1}{B}$ and $r_1(t)\frac{A_1}{B}$, respectively. □

*Proof of Proposition 6.* We define $\Phi(\hat{\varepsilon}, \varepsilon) := u(y(\hat{\varepsilon}), S(\varepsilon))$. Maximizing $\Phi$ pointwise yields the first-order (FOC) and second-order (SOC) optimality conditions at the optimum $\varepsilon = \hat{\varepsilon}$. For notational convenience, we slightly abuse our notation. No confusion should occur. Then,

$$\text{FOC} \quad : \frac{\partial \Phi(\hat{\varepsilon},\varepsilon)}{\partial \hat{\varepsilon}}\Big|_{\varepsilon=\hat{\varepsilon}} = 0 \ ; \qquad \text{SOC} \quad : \frac{\partial^2 \Phi(\hat{\varepsilon},\varepsilon)}{\partial \hat{\varepsilon}^2}\Big|_{\varepsilon=\hat{\varepsilon}} \leq 0.$$

Differentiating the FOC, we get

$$\frac{\partial}{\partial \hat{\varepsilon}}\Big|_{\varepsilon=\hat{\varepsilon}} \left( \frac{\partial \Phi(\hat{\varepsilon},\varepsilon)}{\partial \hat{\varepsilon}}\Big|_{\varepsilon=\hat{\varepsilon}} \right) = \frac{\partial^2 \Phi(\hat{\varepsilon},\varepsilon)}{\partial \hat{\varepsilon}^2} + \frac{\partial^2 \Phi(\hat{\varepsilon},\varepsilon)}{\partial \hat{\varepsilon}\partial \varepsilon} = 0 \Rightarrow \text{SOC} \Leftrightarrow \frac{\partial^2 \Phi(\hat{\varepsilon},\varepsilon)}{\partial \hat{\varepsilon}\partial \varepsilon} \geq 0.$$

Using the explicit expressions for $u(\cdot)$, we get after some calculatoins $\frac{\partial^2 \Phi}{\partial \hat{\varepsilon}\partial \varepsilon}\big|_{\varepsilon=\hat{\varepsilon}} = \int_t^T ds\, e^{-\delta(T-s)} \frac{\partial c_s}{\partial \varepsilon} \frac{\partial^2 u(y(\hat{\varepsilon}),S(\varepsilon))}{\partial \varepsilon \partial c_s}$. Thus, $\frac{\partial^2 \Phi}{\partial \hat{\varepsilon}\partial \varepsilon} \geq 0 \Leftrightarrow \frac{\partial c_s}{\partial \varepsilon} \frac{\partial^2 u(y(\hat{\varepsilon}),S(\varepsilon))}{\partial \varepsilon \partial c_s} \geq 0$. □

*Proof of Proposition 7.* We define the indirect utility function $\hat{U}(\varepsilon) := \max_{\hat{\varepsilon}} \Phi(\hat{\varepsilon},\varepsilon)$. Using the Envelope Theorem we get

$$\frac{d\hat{U}(\varepsilon)}{d\varepsilon} = \frac{\partial u(y(\varepsilon),S(\varepsilon))}{\partial \varepsilon} = \frac{\partial u_1(\hat{c}(\varepsilon),S(\varepsilon))}{\partial \varepsilon}.$$

This implies

$$\hat{U}(\varepsilon) = \underline{\hat{U}} + \int_{\underline{\varepsilon}}^{\varepsilon} \frac{\partial u_1\left(\hat{c}(\tilde{\varepsilon}),S(\tilde{\varepsilon})\right)}{\partial \tilde{\varepsilon}} d\tilde{\varepsilon}.$$

Since the bank maximizes the utility from the joint surplus minus the agent's utility, we get $K(\varepsilon) = \hat{U}(\varepsilon) - v(y(\varepsilon),S(\varepsilon))$, with $v(\cdot)$ defined in (18). The participation constraint (13) then implies $\underline{\hat{U}} = 0$. Hence,

$$\int_t^T e^{-\delta(T-s)} K(\varepsilon) ds = \int_{\underline{\varepsilon}}^{\varepsilon} \frac{\partial u_1\left(\hat{c}(\tilde{\varepsilon}),S(\tilde{\varepsilon})\right)}{\partial \tilde{\varepsilon}} d\tilde{\varepsilon} - u_1(\hat{c}(\varepsilon),S(\varepsilon)).$$

Once we know $c(\varepsilon)$, we know $K(\varepsilon)$. Replacing $K(\varepsilon)$ in the utility function of the bank and carrying out one partial integration, we have proven our claim. □

# Economic capital for market risk

Silvan Ebnöther[*]
Università della Svizzera italiana
and Zurich Cantonal Bank

September 7, 2015

**Abstract**

The interaction of capital and risk for trading and treasury units is of primary interest in the corporate governance of banks as it links operational profitability and strategic risk management. During the financial crisis, several banks' trading units suffered significantly higher losses than their risk capital charged based on value-at-risk constraints. There is a structural inconsistency between strategic risk management with a *one-year* internal capital adequacy assessment process and operating risk management with a *ten-day* risk horizon for trading units.

A new risk budgeting scheme aligns bank risk appetite on an annual basis with operating short-term risk limits. A bank assigns an annual risk budget to its market risk managers. The annual risk budget equals the *probability* of an annual loss that is higher than a predefined capital-at-risk. Managers' risk consumption equals the probability of a year-to-date loss in excess of the capital-at-risk at the 10-day risk horizon. As soon as the risk manager has completely consumed his annual risk budget, he immediately has to hedge his positions. The more risk a manager takes and consumes today, the more likely he will be restricted in the future. This relation forces the risk taker to run market risk in a conscious and far-sighted manner.

*JEL Classification Codes: G32, C51, E22, G21, G31*

**Key words:** Risk management, economic capital, risk capital, value-at-risk, market risk, capital charge, cost of capital, risk budget, exposure management, corporate governance

[*] The author acknowledges Thomas Domenig, who had the idea of consuming probability, and Paolo Vanini for helpful comments and discussions.

## 1 Introduction

The interaction of capital and risk is of primary interest in the corporate governance of banks as it links operational profitability and strategic risk management. [KN92] noted that senior executives understand that their organization's monitoring system strongly affects the behaviour of managers and employees. Typical instruments used by senior executives to focus on strategy are balanced scorecards with objectives for performance and risk management, including an according payroll process. Such a top-down capital-at-risk concept gives the executive board the desired control of the operative behaviour of all risk takers. It guarantees uniform compensation for business risks taken in any division or business area. The standard theory of cost-of-capital (see e.g. [Bas09]) assumes standardised assets. Return distributions are equally normalised to a one-year risk horizon. It must be noted that risk measurement and management for any individual risk factor has a bottom-up design. The typical risk horizon is 10 days for trading positions, 1 month for treasury positions, 1 year for operational risks and even longer for credit risks. In classical theory, one determines capital requirements and risk measurement using a top-down approach, without specifying market and regulation standards. I show how to close the gap between bottom-up risk modelling of short-term market risk and top-down capital alignment.

The financial crisis and financial distresses of several (investment) banks due to large losses by their trading units show that the risk management system in the banking industry failed. There is a conceptual disruption between the top-down enterprise management with economic capital control on an annual cycle and the bottom-up risk taking and limitation system based on a short-term (e.g. ten-day) risk measurement horizon. The common concept of short-term risk constraints such as value-at-risk (VaR) or expected shortfall (ES) does not offer sustainable incentives to manage risk in line with the annual bank risk appetite. Risk guidelines given by pure risk constraints such as VaR limitations are simply fulfilled or not. The risk taker is not punished for taking extra risk as long as he is below the constraint. While an unallocated risk budget is available, the simple and binary VaR-limitation concept does not force risk takers and executive boards to discuss whether additional risk exposure repays enough risk premium. Hence, binary limitation concepts do not sufficiently stimulate any risk culture within a bank. The lack of risk culture was one reason why a financial crisis was possible, see [Kir09], [OEC09] and [BHP09].

A consistent alignment of capital with risk for various units with different associated risk measures is a crucial task. Market risk results from trading and treasury positions. The effective risk horizons for traders and treasurers are much less than one year. It is not reasonable for market risk positions to measure their risk on a one-year horizon due to the implicit assumption that traders or treasurers hold their portfolios constant up to the risk horizon. Since they are continuously adapting and hedging their portfolios due to market shifts, they redeploy their trading positions and treasury exposures several times during a year. Nevertheless, risk capital charges must be based on a single enterprise-wide risk horizon that must not coincide with the risk factor specific horizons. The proposed risk consumption approach brings these different views between (i) risk measurement and management on a short-term horizon and (ii) risk capital on a long-term horizon in line. Whereas market risk has to ”scale up” to a one-year capital management cycle, credit risk has to ”scale down”. [EV07] have shown the alignment of one-year capital with long-term credit risk, proposing a five-year risk horizon for credit portfolio risk modelling.

## 2 Budgeting market risk

Consider a trading unit with a value-at-risk (VaR) on a ten-day horizon and a 99%-confidence level as proposed by the Basel Committee on Banking Supervision ([Bas06]). Its bank allocates capital-at-risk (CaR) based on a one-year horizon and a 99.9%-confidence level, which is in line with the credit risk measurement proposed by Basel II and annual corporate management by objectives.

A simple way to calculate the CaR demands is to scale the VaR limit $\overline{\text{VaR}}$ to the appropriate horizon and confidence level, i.e. $\text{CaR} = \kappa \, \overline{\text{VaR}}$. In the case of a Gaussian setup and 250 bank trading days per year,

$$\kappa = \frac{\sqrt{250}}{\sqrt{10}} \frac{Q\,(99.9\%)}{Q\,(99\%)} = 6.64.^{1}$$

This *VaR scaling* is not satisfactory because it does not take into account aspects of liquidity and hedgeability. The fact that one can hedge and redeploy a trading portfolio in a short time should lead to an advantage in terms of capital charge compared with an illiquid and unhedgeable portfolio. Moreover, a continuous asset reallocation characterises a trading book. Therefore, a holding period of one year is a misplaced assumption.

It becomes evident that a more sophisticated approach is needed to press home the advantage of hedgeability effectively and consistently. If required, a dynamic hedging for risk reduction should be possible. I propose the following budgeting scheme - also illustrated in Figure 1:

- Define (a) the *CaR* [$] and (b) the *confidence level* $1 - \alpha_0$ [%] exogenously at the beginning of a calendar year. $\alpha_0$ defines the *seed budget*. The head of trading chooses at beginning of the year the CaR level needed for his trading strategy. Corporate governance regulates the confidence level and the hurdle rate for the designated CaR.

- At beginning of each consecutive ten-day period during the year, measure the probability $p_t$ that the *cumulative year-to-date loss* at end of period $t$ exceeds the CaR.

- Hedge the portfolio completely as soon as the cumulative probability $\sum_{t=0}^{T} p_t$ exceeds the threshold $\alpha_0$.[2] Maintain the hedge up to the end of the year.

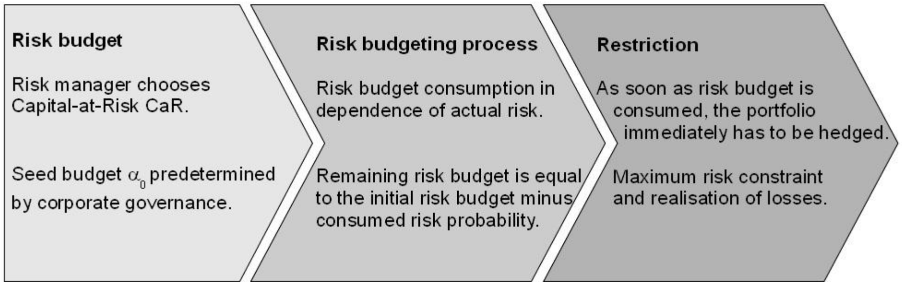

Risk budget
Risk manager chooses Capital-at-Risk CaR.
Seed budget α0 predetermined by corporate governance.
Risk budgeting process
Risk budget consumption in dependence of actual risk.
Remaining risk budget is equal to the initial risk budget minus consumed risk probability.
Restriction
As soon as risk budget is consumed, the portfolio immediately has to be hedged.
Maximum risk constraint and realisation of losses.

Figure 1: Risk budgeting scheme

Formal definitions for the scheme can be found in Chapter 3. I further extend the approach to a continuous time framework in Appendix 5.2.

The budgeting scheme is complementary to the VaR approach in many aspects. First and most important, a high level of risk consumes more budget than a low level, even if there are no realised losses. The budgeting is path-dependent both with respect to realised profit and loss (P&L ) and realised past exposures. A positive P&L decreases the *marginal* budget consumption at a constant level of risk in a non-linear manner, and vice versa. Year-to-date profits offer more flexibility in risk taking to the risk manager. Year-to-date losses decrease the distance-to-CaR and increase the risk consumption per unit of risk exposure taken. The risk manager has to continuously reduce his positions to prevent a complete constraint. The budgeting is essentially a stop-loss mechanism. In the case of complete liquidity, i.e. the portfolio could be hedged instantaneously, the budgeting would collapse to a simple stop-loss order, i.e. to liquidate the portfolio as soon as the capital is consumed. Note that the VaR scaling approach does not have a built-in stop-loss. A major drawback of conventional VaR limitations is that no continuous and foresighted risk strategy is needed. A simple point-in-time risk management satisfies any risk requirements. The risk budgeting scheme closes the gap since todays risk impacts the remaining risk budget and future flexibility.

Second, high autocorrelation in underlying movements and volatilities during market downturns increases the risk of ”clustered” losses. Pure VaR

[1] $Q(x)$ defines the quantile of the standard normal distribution at the $x\%$-confidence level. ($Q(99\%) = 2.33$; $Q(99.9\%) = 3.09$)

[2] I approximate the cumulative probability with the sum of the exceeding probabilities $\sum_{t=0}^{T} p_t$ instead of using the product rule $1 - \prod_{t=0}^{T}(1 - p_t)$, since (i) it is a simplification and (ii) the approximation error is negligible for probabilities less then 0.1%.

approaches neither avoid nor punish this empirical observation. The budgeting scheme directly incorporates (clustered) losses in the budget consumption and prospective risk capability. It ensures that the ruin probability, conditioned on the profit-and-loss and volatility realisation, is in line with the CaR assumptions at any time. It does so by effectively transforming the tail of the one-year loss distribution by imposing path-dependent hedging constraints.

Another distinction between the two approaches is the strong *seasonality* inherent in the budgeting approach. Seasonality arises naturally with the stepwise expiration of the annual risk budgeting period. If a risk manager has consumed less risk budget through the year than assumed at the time of application, he can take significantly more risk exposure for the rest of the year. One avoids seasonality by defining a parallel budgeting on two or even more rolling one-year windows. This enlargement is more expensive to handle due to parallel management of several evolving budgets. I discuss some aspects of seasonality in Chapter 4.3.

The budgeting scheme delegates the investment decision about (i) point in time and (ii) level of budget consumption to the risk manager. As long as the risk manager feels comfortable with the budget consumption, he is free to hold ”high-risk” positions temporarily even though his VaR is above target and he disproportionately consumes risk budget. Needless to say, he also has to reduce his risk with a budgeting scheme in the medium term, but not immediately. An immediate closing of significant positions during a stressed and probably illiquid market is (often) very expansive. Such strategies and risk concepts compromise the bank and the market. The risk manager probably wins a bit more time. On the other hand, he will be punished progressively to the end of the year with less remaining risk budget.[3] This behaviour is in line with the incentives to govern risk readiness on an annual basis.

[3] If corporate governance demands a risk cap, it is straightforward to define an additional and conventional VaR constraint.

## 3 Definitions and implementation

The concept of the sketched budgeting approach includes drawbacks and opportunities. Since there are no assumptions concerning the portfolio evolution, and since risk only has to be measured on a short time horizon, errors caused by misspecification are smaller compared with approaches that measure risk on a one-year horizon and with predefined portfolio dynamics[4].

The key question concerning the modelling lies in the tail assumption of the profit & loss (P&L ) distribution. $l_t$ denotes the P&L in period $t$, $L_t := \sum_{s=0}^{t} l_s$ the cumulative year-to-date P&L at end of period $t$ and $D_t = \mathrm{CaR} + L_{t-1}$ the distance-to-CaR at beginning of period $t$. The cumulative distribution function corresponding to the CaR quantile specifies the budget consumption $p_t$ in period $t$ as follows:

$$p_t := P[L_t < -\mathrm{CaR}|\mathcal{F}_{t-1}] = P[l_t < -D_t|\mathcal{F}_{t-1}] \qquad (1)$$

See Figure 2.

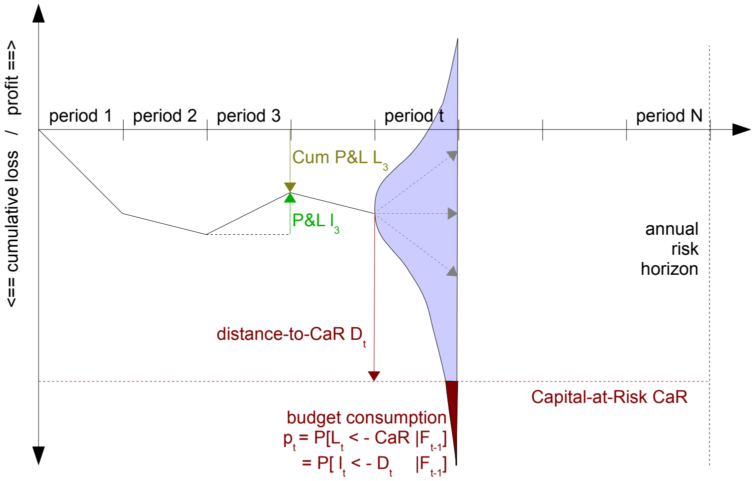


Figure 2: The marginal budget consumption $p_t$ equals the probability of a year-to-date loss in excess of the capital-at-risk.

[4] One could probably also assume a constant portfolio.

Considering an ordinary trading portfolio, a variety of risk factors and non-linear payoff profiles of (exotic) derivatives impact the actual loss distribution, which is accordingly complex and analytically not tractable.[5] Moreover, for calculating the marginal budget consumption, one has to compute "very high" quantiles far beyond quantile $\alpha_0$ (i.e. $\alpha_0 = 10$ bp) representing the seed budget. Recall that in the case of 25 non-overlapping ten-day periods in a year, the average marginal consumption should be less than $\alpha_0/25$ (0.4 bp) on average to grasp relevant quantiles. The crucial part of the distribution reaches from less than $\alpha_0/100$ (0.1 bp) to perhaps $\alpha_0$ (10 bp).[6] The residual body of the ten-day distribution is negligible. This fact holds for most one-period risk measures, too. Of course, statistical assumptions about the distribution of such high quantiles are speculations. The relevant quantiles are far in excess of levels for which one can calculate statistically significant statements on realised losses. An estimate of a distribution based on a bigger Monte-Carlo simulated loss sample is also not reasonable since model risk is huge in the tails. I therefore assume an analytic distribution with desired properties. I choose the family of normal inverse Gaussian distributions (NIG) and their associated Lévy processes. Another "consensus distribution" for ten-day returns might be a student-t distribution with six degrees of freedom.[7] The family of NIG distributions is better suited than the family of t-distributions from the point of view that the NIG family is invariant under convolution and it is infinitely divisible.

## 3.1 Understanding NIG distribution

The NIG distribution depends on four parameters, namely $\mu$ (location), $\beta$ (skewness), $\alpha$ (tail heaviness) and $\delta$ (scale). There are two restrictions, namely $0 \leq |\beta| < \alpha$ and $\delta > 0$. There is an analytic density for the NIG distribution, and simple simulation methods are available. Some properties are the following:

[5] All discussions in the article are based on the assumption that the underlying portfolio is characteristic for a bank with a universal trading unit.

[6] The case of 25 non-overlapping periods is the simplest implementation of the scheme. However, the crucial distribution element remains the same, also if the scheme is enhanced to 252 overlapping periods.

[7] When choosing the parameters, one should keep in mind that the tail fatness depends on the chosen process for the underlying volatility. In a GARCH type or stochastic volatility model, for instance, the tail of the unconditional distribution will usually be fatter than that of the conditional distribution given the volatility, so that a slightly less fat-tailed distribution for the innovation process might be appropriate.

| horizon | 1d | **10d** | 1m | 2m | 3m | 6m | 12m |
|---|---|---|---|---|---|---|---|
| $\delta\alpha$ | 0.1 | **1** | 2.083 | 4.167 | 6.25 | 12.5 | 25 |
| (excess) kurtosis | 30 | **3** | 1.44 | 0.72 | 0.48 | 0.24 | 0.12 |

Table 1: Resulting kurtosis as a function of risk horizon assuming a kurtosis of 3 at the ten-day horizon.

- The scale property is

$$X \sim \mathcal{NIG}(\mu, \alpha, \beta, \delta) \Longrightarrow cX \sim \mathcal{NIG}(c\mu, \alpha/c, \beta/c, c\delta) \tag{2}$$

  If $\mathcal{NIG}(\mu, \alpha, \beta, \delta)$ is the distribution of a NIG process at time $\bar{t} = 1$, the distribution for all $t > 0$ is given by

$$\mathcal{NIG}(t\mu, \alpha, \beta, t\delta). \tag{3}$$

  Generally, for a Lévy process, $X_t$ with $X_1 - X_0 \sim \mathcal{NIG}(\mu, \alpha, \beta, \delta)$ results in

$$\frac{X_t - X_0}{\sqrt{t}} \sim \mathcal{NIG}(\sqrt{t}\mu, \sqrt{t}\alpha, \sqrt{t}\beta, \sqrt{t}\delta). \tag{4}$$

- The parameters define the first four moments. I define $\gamma = \sqrt{\alpha^2 - \beta^2}$ such that the mean is $\mu + \frac{\delta\beta}{\gamma}$, the variance $\frac{\delta\alpha^2}{\gamma^3}$, the skewness $\frac{3\beta}{\alpha\sqrt{\delta\gamma}}$ and the (excess) kurtosis $\frac{3}{\delta\gamma}\left(1 + \frac{4\beta^2}{\alpha^2}\right)$.

I refer to [BN97a], [BN97b], and [VdJ02] for details concerning this class of distributions and processes.

Since extreme losses and their tail distribution are the factors of interest, I constrain the distribution parameters to symmetric NIG distributions. The mean and the skewness are zero if $\mu = \beta = 0$. It follows variance $\frac{\delta}{\alpha}$ and kurtosis $\frac{3}{\delta\alpha}$. I calibrate the variance to the VaR later. I normalise the distribution so that variance $\frac{\delta}{\alpha}$ is 1, i.e. $\alpha = \delta$. I assume a "medium" kurtosis of 3 at a ten-day horizon.[8] Table 1 shows kurtosis with respect to different risk horizon. Empirical values for short horizons, cited in [Ryd97] and [KSW07] and [VdJ02], show that my assumption is rather conservative. In particular the kurtosis of the one-day distribution seems overvalued. However, the studies analyse stock return distributions, while I consider the distribution of trading portfolios with (exotic) derivatives on arbitrary underlying assets.

[8] The student-t distribution with six degrees of freedom has a kurtosis of 3, too.

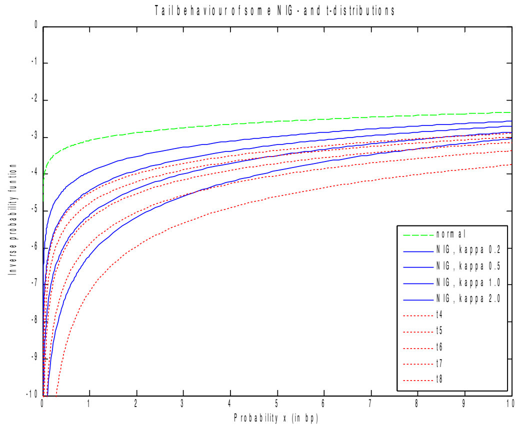


Figure 3: Tail behaviour of different distribution functions.

The nonlinearity of trading books, which contain options, leads to heavier tails than in the pure equity portfolios. Kurtosis must therefore be robustly estimated on any existing trading book. The financial crisis demonstrates that a conservative risk policy benefits in the long run.

I chose the normalised distribution $F_{NIG(\delta,\alpha)}$. I expect to have an accurate estimation $\hat{\text{VaR}}$ of VaR at the 1% quantile.[9] I define the P&L distribution per period by linear scaling of the normalised distribution by factor

$$\zeta = \frac{\hat{\text{VaR}}}{F_{NIG(\delta,\alpha)}(1-\rho_{VaR})}.$$

One might ask what the consequences and model risks of this type of modelling choice are. For this, I compare the tail behaviour of several NIG distributions and t-distributions in Figure 3. I show that the NIG quantile for a ten-day horizon is approximately the quantile of the t distribution with five degrees of freedom, which itself is often called by experts of trading risks. I justify this perhaps rather conservative choice of NIG distribution by the fact that the distribution should also include event and default risk.

[9] Risk portfolio systems model the loss distribution of a trading portfolio using Monte Carlo methods with tens of thousands of simulations. These methods allow quite reliable estimations of quantiles on a ten-day horizon.

### 3.2 Example I

An example illustrates the risk budgeting procedure. The following setup is used: (i) I divide the year into 25 ten-day periods. (ii) The CaR equals USD 50 m. (iii) The seed budget $\alpha_0$ equals $10bp$. (iv) The probability of a cumulative loss bigger than CaR follows the NIG distribution with $\alpha = \delta = 1$, scaled so that a 99% VaR of USD 10 m results. (v) The realised P&L per period ($l_t$ for $t \in \{1, ..., N\}$) is USD -4 m.

As is shown in table 2, the risk manager must reduce his positions in a risk point of view prior to period 10, otherwise he would consume more budget (23.3032) than he has left (2.0298). He has to account some risk budgets for all pending periods to the end of the year. I present different aspects of a skilled risk strategy in Chapter 4. I also mention a continuous version of the budgeting scheme in Chapter 5.2.

| Period t | distance to CaR | consumption | risk budget | P&L | P&L YtD |
|---|---|---|---|---|---|
| | $D_t$ | $p_t$ | $\alpha_t$ | $l_t$ | $L_t$ |
| | [m] | [bp] | [bp] | [m] | [m] |
| Period 1 | 50 | 0.0003 | 9.9997 | -4 | -4 |
| Period 2 | 46 | 0.0009 | 9.9989 | -4 | -8 |
| Period 3 | 42 | 0.0029 | 9.9959 | -4 | -12 |
| Period 4 | 38 | 0.0099 | 9.9860 | -4 | -16 |
| Period 5 | 34 | 0.0339 | 9.9521 | -4 | -20 |
| Period 6 | 30 | 0.1183 | 9.8338 | -4 | -24 |
| Period 7 | 26 | 0.4211 | 9.4127 | -4 | -28 |
| Period 8 | 22 | 1.5404 | 7.8724 | -4 | -32 |
| Period 9 | 18 | 5.8425 | 2.0298 | -4 | -36 |
| Period 10 | 14 | 23.3032 | consumed | 0* | -36 |
| Period 11-25 | 14 | - | consumed | 0* | -36 |

Table 2: Example I: Risk budgeting in a complete liquidity setup. (*) Since trading and treasury portfolios are not completely hedgeable in practice, a realised P&L and VaR of zero will not be realistic.

# 4 Exposure management and incentives

## 4.1 Choosing capital-at-risk

The executive management reviews the economic capital at beginning of each year. The top-down question for management is to define a suitable capital-at-risk on a 0.1% confidence level. The bottom-up challenge for management is supplying enough risk capital to the trading unit that the head of trading can define appropriate risk limits (e.g. VaR limits) for his trading desks such that they can achieve their objectives. For choosing the CaR itself, one can use the same indicators and sensitivities as for the exposure management during the budgeting process. When choosing the CaR, one should keep in mind that it refers to a very high confidence level and one should be able to absorb huge P&L shocks.

A helpful multi-period indicator is restriction probability $R_t$ defined as the probability that the budget will be consumed by the end of the year, assuming constant value-at-risk and that the P&L distribution equals the tail distribution.[10]

$$R_t = \mathbb{P}\left\{ \bigcup_{\tau > t} \alpha_\tau < 0 | X_\tau - X_{\tau-1} \sim X_t - X_{t-1} \quad \forall \tau > t; \quad \mathcal{F}_{t-1} \right\} \tag{5}$$

The CaR is defined implicitly such that restriction probability corresponds to the seed risk budget ($R_0 = \alpha_0 = 10$bp). In doing so, the CaR Level depends on the tail assumption of the P&L distribution. Therefore, choose an adequate and constant VaR that equals the average VaR taken by the risk manager during the year. The corresponding CaR for Example 1 (i)-(iv)[11] is USD 80.9 m. This is higher than if it were calculated using the pure VaR-scaling approach. ($\kappa = 6.64 \Rightarrow \tilde{CaR} = \text{USD } 66.4m$). Figure 4 compares the one-year cumulative P&L with and without the stop-loss of the risk budgeting approach using assumptions (i)-(iv) of Example 1. Example 2 of Appendix 5.3 also illustrates the impact of the stop-loss order.

Another helpful analysis is to find the implicit and path dependent random number $\overline{CaR}$ such that the risk budget is exactly consumed by the end

[10] This assumption reduces complexity. Since we have made the NIG assumption based on tail properties, another distribution could better match the main body of real distribution. One can estimate this ”second” distribution to empirical or MC-simulated P&L figures, in contrast to the tail estimation.

[11] 99%-VaR = USD 10 m

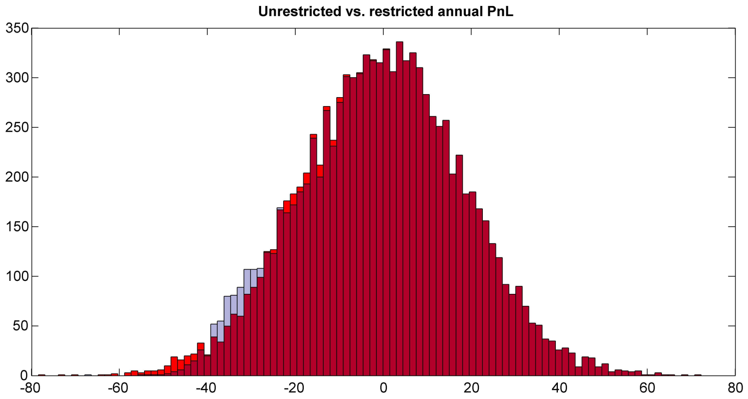


Figure 4: Cumulative P&L with and without stop-loss of the risk budgeting. Note that the risk budgeting cuts the distribution on the left-hand side (losses), although the right-hand side of the distribution (profits) is hardly affected.

of the year. Figure 5 shows the histogram of $\overline{CaR}$ with the assumption of Example 1 (i)-(iv). Notice that the 99.9% quantile of $\overline{CaR}$ is again USD 80.9 m.

## 4.2 Adequate risk taking and local utility

A key question for risk managers is to know the level of exposure (say VaR) that is in line with the risk budget and capital restrictions. He needs to know at any time (i) how much risk he may enter into and (ii) the likelihood of a restriction. He also needs some indications about any necessary or potential risk reduction that reduces his risk budget consumption and restriction probability.

Since the marginal consumption is highly non-linear and increases rapidly in the event of a P&L shock, the remaining budget itself is a poor indicator. Although risk and loss per period are constant in Example 1, risk consumption increased exponentially after period 6 - as seen in Table 2. Stable, adequate risk-taking usually leads to partial budget consumption, i.e. the vast majority of the risk budget is in surplus at year-end.

A better control variate than the remaining risk budget is local utility

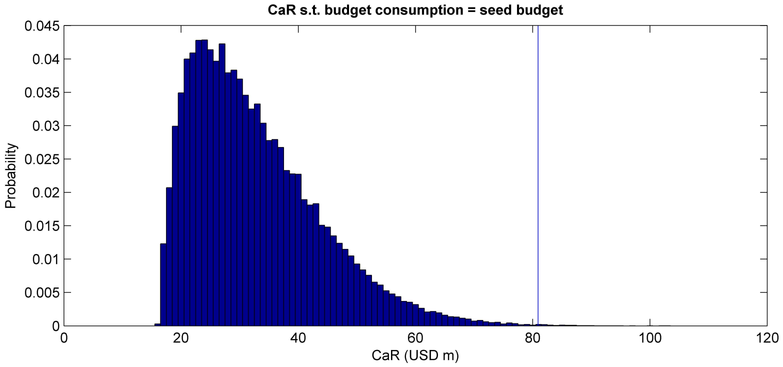


Figure 5: Crucial CaR: path dependent CaR such that risk budget is exactly consumed by the end of the year.

$u_t$, defined as

$$u_t := \frac{F^{-1}_{NIG(\delta,\alpha)}\left(\frac{\alpha_{t-1}}{N-(t-1)}\right)}{D_t}, \quad \forall\, t \in \{1, ..., N\}.$$

$F^{-1}_{NIG(\delta,\alpha)}(x)$ is the inverse distribution function of a NIG distribution. Notice that $\delta$ and $\alpha$ are both derivatives of the assumed VaR. A local utility of 1.0 means that the proportionally allocated part of the remaining risk has been used accurately in the current budget period. As long as the manager complies with a local utility of less than 1.0, he is not restricted.

## 4.3 Seasonality

The adaptive exposure strategy represents quite an unorthodox and perhaps impractical way of managing risk. It fully exploits the strong seasonality in the budgeting, i.e. the positive theta effect of the risk acceptance level. In practice, an almost unlimited increase of risk exposure towards the end of the year would be objectionable and undesirable. Even if the capital allocation process is annual, operating profits and losses are communicated more than once a year. It is therefore undesirable to achieve large cumulated profits until the middle of December and then lose it all in a fortnight. One can avoid such a scenario by imposing a subsidiary ten-day value-at-risk limit.[12] I reject this extension because it is conceptually inconsistent and

[12] Being only of subsidiary character, such a limit can be allocated generously. Also, it will not induce any capital costs, as these are already being taken care of by the CaR limit.

objectionable.

A second and better approach to avoid seasonality is a parallel budgeting on two or even more rolling one-year windows. The residual risk budget $\alpha_{t_k}$ reads

$$\begin{aligned}
\alpha_t &= \alpha_0 - \sup_{T \in \{t-N,t\}} \sum_{\tau=T}^{t} p_\tau^T, && (6)\\
p_\tau^T &= \mathbb{P}\left(l_\tau \leq -D_\tau^T | \mathcal{F}_{\tau-1}\right), && (7)\\
D_\tau^T &= \mathrm{CaR} - \sum_{\bar{t}=T}^{\tau} l_{\bar{t}} && (8)
\end{aligned}$$

where $l_{\bar{t}}$ is the P&L in period $\bar{t}$. Due to the path dependency of marginal consumption, implicit parameter $T$ of equation 6 is not always equal to $t - 25$. This enlargement would be difficult to handle since it would imply managing up to 25 evolving budgets at the same time. A parallel budgeting on a monthly or quarterly basis is enough to obviate undesirable seasonality effects.

### 4.4 Handling of restrictions

The crucial point of the budgeting scheme is to interpret the restriction event in practice and take reasonable steps. The budgeting concept stipulates that the risk manager has to hedge his positions after he has exhausted the risk budget and no further trading activity is foreseen for rest of the year. A rigorous application of the budgeting scheme does not allow any tolerance. Hence, the manager cannot apply for additional risk capital. The consequences could be too drastic, irrational and non-enforceable in practice. However, it is a fundamental misconception that the budgeting scheme is a free lunch on additional risk capital. The fundamental idea of the concept is to align short-time risk with annual risk capital and to achieve comparability over different risk categories - including short-term market risk and long-term credit risk.

Since the complete outstanding of any trading activity is inappropriate sometimes, the bank's top management would perhaps accept an exceptional request for new risk capital in compliance with rigorous conditions and consequences such as the dismissal of the risk manager.

# 5 Appendix

## 5.1 Notations

| parameter | meaning | example |
|---|---|---|
| CaR | capital-at-risk | US 100 m |
| VaR | value-at-risk at quantile $\rho_{VaR}$ | USD 20 m |
| $\rho_{VaR}$ | quantile of VaR measure | 99% |
| H | risk horizon | 10 business days |
| T | budget horizon | 1 year |
| N | number of periods per year | 25 |
| t | period $t$ | $t \in \{1, ..., N\}$ |
| $l_t$ | profit & loss for period $t$ | USD - 5 m |
| $L_t$ | cumulative profit & loss for periods 1 to $t$ | USD -30 m |
| $D_t$ | distance-to-CaR at beginning of period $t$ | USD 70 m |
| $\alpha_0$ | seed budget | 10 bp |
| $\alpha_t$ | risk budget at end of period $t$ | 8 bp |
| $p_t$ | budget consumption in period $t$ | 2 bp |
| $u_t$ | local utilisation in period $t$ | 50% |
| $R_t$ | restriction probability at period $t$ | 10 bp |
| $F_{NIG}(x)$ | cumulative distribution function for NIG distribution | $F_{NIG}(0.0) = 0.5$ |
| $F_{NIG}^{-1}(x)$ | inverse distribution function for NIG distribution | $F_{NIG}^{-1}(0.5) = 0.0$ |
| $Q(\alpha)$ | $\alpha$ quantile (of the standard normal distribution) | $Q(99.9\%) = 3.09$ |

## 5.2 Risk budgeting in continuous time

I have presented the budgeting approach in a discrete setup due to practicability. I note that a continuous version of the residual risk budget $\alpha_t$ could be formally written as

$$\begin{aligned} \alpha_t &= \alpha_0 - \bar{\gamma} \int_0^t p_\tau d\tau \\ p_\tau &= \mathbb{P}\left( L_{\tau+10d} \leq -CaR | \mathcal{F}_\tau \right), \end{aligned}$$

where $\bar{\gamma} \approx \gamma(\tau)$ equals 1 over the year fraction between $\tau$ and the relevant risk horizon ($\tau + 10$ business days), i.e. $\bar{\gamma} \approx 252/10$. One can also extend the continuous setup with the concept of parallel budgeting to avoid seasonality. The residual risk budget $\alpha_t$ reads

$$\begin{aligned} \alpha_t &= \alpha_0 - \gamma \sup_{T \in \{t-1, t\}} \int_T^t p(T, \tau) d\tau, \\ p(T, \tau) &= \mathbb{P}\left( L(\tau, \tau + 10d) \leq -D(T, \tau) | \mathcal{F}_\tau \right), \\ D(T, \tau) &= \mathrm{CaR} - L(T, \tau), \end{aligned}$$

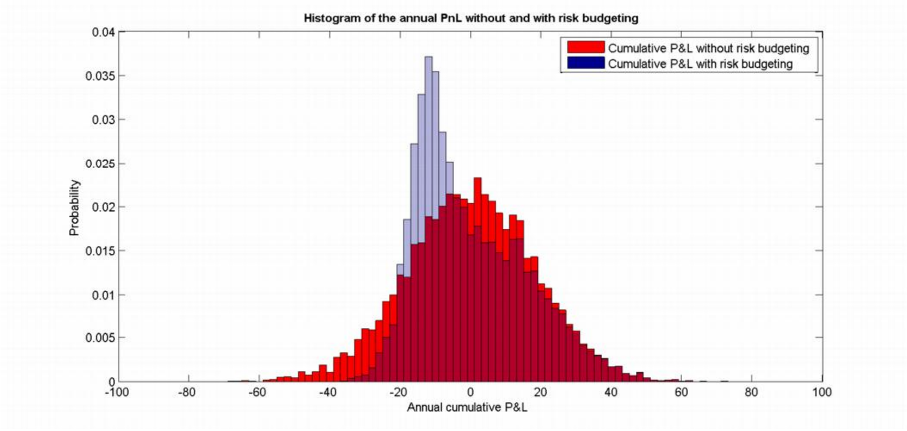


Figure 6: Cumulative P&L with and without risk budgeting for Example 2. Note that the risk budgeting cuts the distribution on the left-hand side (losses), although the right-hand side of the distribution (profits) is hardly affected.

where $L(T, \tau)$ is the cumulative profit & loss between $T$ and $\tau$.

The implementation of the continuous version will fail in practice since reproducibility of all risk figures is a precondition of corporate governance. Moreover, the computational requirements for continuous intraday risk measurement of a trading portfolio including exotic options are too huge, also in the foreseeable future.

A feasible adaptation to the presented approach with discrete periods is a daily risk measurement and risk budgeting. The budgeting scheme described in chapter 2 remains in place except that the daily risk consumption equals the tenth part of the probability of a cumulative year-to-date loss exceeding the CaR at the ten-day risk horizon.

### 5.3 Example 2

Example 2 corresponds to Example 1 except that the CaR is USD 30 m. Because of the high restriction probability $R_0 = 50.8\%$, the consequence of the risk budgeting is more evident than in Example 1. Figure 6 compares the one-year cumulative P&L with and without the risk budgeting approach.

## References

[ABA02] Francesco Audrino and Giovanni Barone-Adesi. A multivariate fgd technique to improve var compuatition in equity markets. *FinRisk*, 2002.

[Bas06] Basel Committee on Banking Supervision. International convergence of capital measurement and capital standards: A revised framework - comprehensive version. June 2006.

[Bas09] Basel Committee on Banking Supervision. Range of practices and issues in economic capital modeling. 2009.

[BHP09] H.J. Blommestein, L. H. Hoogduin, and J. J. W. Peeters. Uncertainty and risk management after the great moderation: The role of risk (mis)management by financial institutions. *Paper for the 28th SUERF Colloquium on 'The quest for stability', Utrecht, The Netherlands*, 2009.

[BN97a] Ole E. Barndorff-Nielsen. Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24(1):1–13, March 1997.

[BN97b] Ole E. Barndorff-Nielsen. Processes of normal inverse gaussian type. *Finance and Stochastics*, 2(1):41–68, 1997.

[EV07] Silvan Ebnoether and Paolo Vanini. Credit portfolios: What defines risk horizons and risk measurement? *Journal of Banking and Finance*, 31(12):3663–3679, 2007.

[GTV10] Leo Grepin, Jonathan Tétrault, and Greg Vainberg. After black swans and red ink: How institutional investors can rethink risk management. *McKinsey Working Papers on Risk*, (17), 2010.

[Kir09] Grant Kirkpatrick. Corporate governance lessons from the financial crisis. *Financial Market Trends*, 1(96), 2009.

[KN92] Robert S. Kaplan and David P. Norton. The balanced scorecard œ measures that drive performance. *Harvard Business Review*, January-February 1992.

[KSW07] Anna Kalemanova, Bernd Schmid, and Ralf Werner. The normal inverse gaussian distribution for synthetic cdo pricing. *Journal of Derivatives*, 14(3), Spring 2007.

[OEC09] OECD. Corporate governance and the financial crisis: Key findings and main massages. June 2009.

[Ryd97] T.H. Rydberg. The normal inverse gaussian lévy process: simulation and approximation. *Communications in Statistics: Stochastic Models*, 13(4):887–910, 1997.

[Sai04] Francesco Saita. Risk Capital Aggregation: The Risk Manager's Perspective. *EFMA 2004 Basel Meetings Paper*, 2004.

[VdJ02] Johannes H. Venter and Pieter J. de Jongh. Risk estimation using the normal inverse gaussian distribution. *Journal of Risk*, 2:1–25, 2002.