ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS (ESE)
MASTER THESIS FINANCIAL ECONOMICS

On the Value-at-Risk of banks

- A European perspective -

AUTHOR	SANDER VAN DEN BOOGAART
STUDENT NUMBER	294831
SUPERVISOR	DR. R.C.J. ZWINKELS
THIS VERSION	October 17, 2009

On the Value-at-Risk of banks: a European perspective

SANDER VAN DEN BOOGAART*

SUPERVISOR:

DR. R.C.J. ZWINKELS

Erasmus University Rotterdam

Erasmus School of Economics (ESE)

Master Thesis Financial Economics

This version: October 17, 2009

Abstract

Risk management can be considered as one of the core activities of banks. Especially in the current financial crisis, sound risk management is of the utmost importance. The Basel Committee on Banking Supervision has played a central role in the achievement of a regulatory framework to secure proper risk management of internationally active banks. One of the aspects of this framework concerns the treatment of market risk. To measure market risk, banks are allowed to use internal Value-at-Risk models. We extend the field of empirical research with an analysis of the quality of these models, using a sample of five large European banks. Furthermore, the banks' internal models are compared with five benchmark models. We find that four out of five internal models are incorrectly specified; three banks overstate their risks and one bank understates its risks. Also, several benchmark models (e.g. a RiskMetrics model) yield better Value-at-Risk estimates than the internal models. The degree of risk under- or overstatement is quantified and the costs of overstatement are expressed in monetary terms. The results are mixed, ranging from an understatement of 6,5% to an overstatement of 63,1%. Monetary costs of risk overstatement tend to be rather small in the context of the banks' financial statements.

* Student number: 294831

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

At the moment of writing this thesis the world is struck by an unprecedented global financial crisis. Up to now, several (large) banks and other financial institution collapsed under the burden of the crisis. This has led people to question the quality of risk management and the effectiveness of regulation in the field.

Risk management can be considered as one of the core activities of banks. Understanding risks and handle them appropriately is crucial for every bank in order to survive in the market environment. Especially in the current financial crisis, sound risk management is of the utmost importance.

Over the years, the measurement and management of financial risks has also become the subject of attention of regulatory bodies. The Basel Committee on Banking Supervision has played a central role in the achievement of a regulatory framework to secure proper risk management of internationally active banks. In its revised framework the Basel Committee (2006, p 2.) states that its fundamental objective “has been to develop a framework that would further strengthen the soundness and stability of the international banking system while maintaining sufficient consistency that capital adequacy regulation will not be a significant source of competitive inequality among internationally active banks”.

One of the aspects of the regulatory framework concerns the treatment of market risk, which arises from the volatility in the trading book of banks. To measure market risk, banks are allowed to use internal risk models. These models are based on the Value-at-Risk metric (VaR) -

the maximum potential loss of a position within a given time frame and confidence level - which is a well known measure in the international world of finance. The quality of these internal Value-at-Risk models of banks is validated on an ongoing basis by the supervisory authorities; this is called backtesting. In practice, this will come down to an ex post comparison of the daily Value-at-Risk measures of a bank with the subsequent daily profit or loss arising from the trading book.[1] Based on these VaR estimates, banks have to meet certain capital requirements.

A vast amount of papers in the academic field has been dedicated to the development and the examination of the validity of different types of Value-at-Risk models.[2] In another, related literature stream different backtesting procedures are reviewed or proposed.[3] However, not much is known about the type and quality of Value-at-Risk models as applied by banks. This may be explained by the fact that the bank models for calculating market risks are not public. Also, the daily profit or loss and VaR estimates are not disclosed.

Only recently, researchers have shifted their attention to the empirical study of Value-at-Risk models. Berkowitz and O'Brien (2002) were the first to use empirical data to examine the quality of the VaR estimates of banks. They used (anonymous) data of six United States commercial banks to perform their investigation. The authors find that banks overstate their risks. Furthermore, the internal banks' models do not provide better estimates of Value-at-Risk than a simple ARMA(1,1)-GARCH(1,1) model. Pérignon, Deng and Wang (2008) developed a method to extract data concerning profits and losses and VaR estimates from the annual reports of banks. By applying this method to the annual reports of six Canadian banks, they were the first with non-anonymous data at their disposal. The findings of Pérignon *et al.* (2008) are in line with those of Berkowitz and O'Brien (2002); risk estimates are too conservative. The degree of risk overstatement is quantified and varies from 19% to 79%.

The goal of this thesis is to extend the field of empirical research with an analysis of the quality of the VaR figures reported by five large European banks. For this purpose, the methodology applied in the thesis will partially be based on the work of Pérignon *et al.* (2008). The empirical part of the thesis will consist of a comparison between the Value-at-Risk figures of the five banks and a number of benchmark models based on the profit and loss series of these

[1] Throughout this thesis, we refer to the profit or loss arising from the trading book as 'trading revenue', 'P&L' or 'profit and loss series'. Thus, these terms are used interchangeably.

[2] See, for example, Jorion (2007) for an overview of Value-at-Risk models in finance.

[3] See, for example, Lopez (1999) or Campbell (2005) for an overview of backtesting approaches.

banks. The benchmark models include two models based on historical simulation, an ARMA(1,1)-GARCH(1,1) model and the RiskMetrics model. Coverage tests will be executed to compare the performance of the different models. The research continues with an analysis of the amount of risk understatement or overstatement if it turns out to be the case that the models are not correctly calibrated. Pérignon *et al.* (2008) use a so-called 'risk deviation coefficient' to measure risk *overstatement*, which can take on positive values between zero and one. However, based on our expectations, we replace the risk *overstatement* coefficient with a risk *deviation* coefficient, which can take on *both* positive and negative values. Through this adjustment it is possible to also quantify the degree of risk *understatement*.

In short, we find that four out of five bank models are rejected, based solely on unconditional coverage tests. Thus, these models are too tight or too conservative. If we take the aforementioned benchmark models into account, then we find that some of these models - especially the 250-day historical simulation model and the RiskMetrics model - perform better than the banks' internal models. Furthermore, our findings indicate that three banks overstate their risks, one bank understates its risks and only one bank has a correctly specified model.

After this rough picture of what can be expected in this thesis, the lay-out of the remainder is as follows. First of all, the theoretical framework will be covered in chapter 2. Attention will be paid to the Basel regulation and the existing literature in the field. In chapter 3, the choice for the data and the methodology will be justified. Also, the data extraction method developed by Pérignon *et al.* (2008) will be reviewed. In chapter 4, the results of the empirical research will be discussed. In the conclusion, the findings are reviewed and some final remarks will be made. In particular, the results are compared with those of prior empirical research.

CHAPTER TWO

BACKGROUND

This chapter is devoted to a description of the regulatory framework controlling the banking industry. Furthermore, a concise but complete overview of the literature in the field of Value-at-Risk and of the empirical studies of banks' internal market risk models will be presented.

2.1 THE BASEL FRAMEWORK

The Basel Committee on Banking Supervision, established in 1974, is an institution that consists of banking supervisory authorities. The Committee has been founded by the central bank governors of the Group of Ten (G-10) countries.[4] The purpose of the Committee is not to force its member states to adopt the rules it creates, but to formulate broad supervisory guidelines and standards and to recommend statements of best practice. Thus, the regulatory framework developed by the Committee over the years does not have any legal power. However, the Committee does expect that member countries implement the regulatory framework in their respective legislations. Over the years several documents have been released by the Committee.

The Basel Capital Accord of 1988 (also known as 'Basel I') can be considered as the first major set of standardized rules and has been adopted by supervisory authorities all over the world. In this framework two "fundamental objectives" were formulated by the Committee. Paragraph 3 of the Accord states that "these are, firstly, that the new framework should serve to strengthen the soundness and stability of the international banking system; and, secondly, that

[4] As of June 10, 2009, the Committee consists of representatives of the following members: Argentina, Australia, Belgium, Brazil, Canada, China, France, Germany, Hong Kong, India, Indonesia, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, Russia, Saudi Arabia, Singapore, South Africa, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States.

the framework should be fair and have a high degree of consistency in its application to banks in different countries with a view to diminishing an existing source of competitive inequality among international banks".

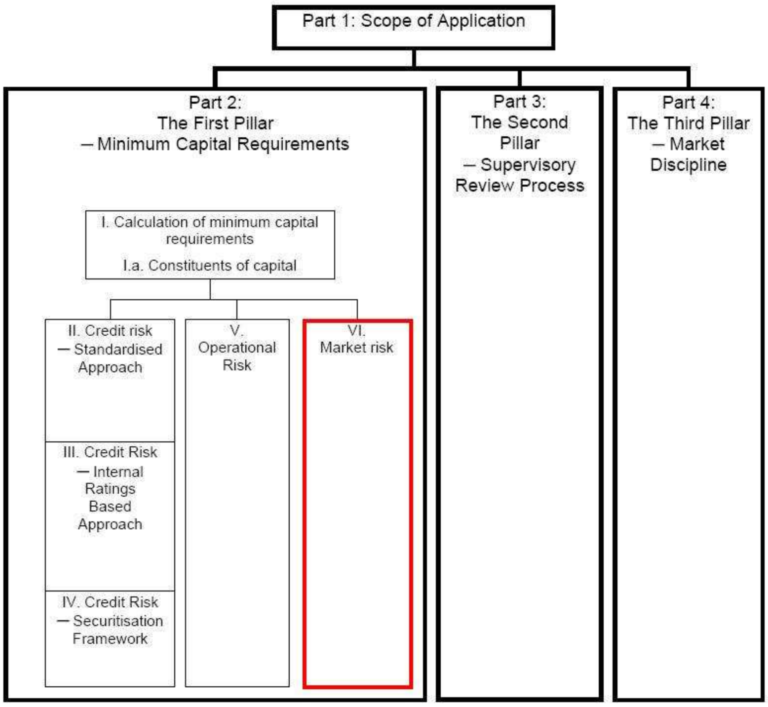
The Accord of 1988 primarily aimed at minimum capital requirements to limit credit risk. Minimum capital requirements with respect to market risk arising from the trading positions of a bank were not yet covered by this framework. Therefore, in 1996 the Amendment to the Capital Accord to incorporate market risks was published. This document was supposed to fill the blank that was left open by the 1988 Accord. More formally, in paragraph 1 of the Amendment market risk is defined as "the risk of losses in on- and off-balance sheet positions arising from movements in market prices". Two types of risks can be separated: (1) "the risks pertaining to interest rate related instruments and equities in the trading book" and (2) "foreign exchange risk and commodities risk throughout the bank".

In June 2004 a revised capital framework was issued by the Committee, the International Convergence of Capital Measurement and Capital Standards: A Revised Framework (also known as 'Basel II'). A comprehensive document was published in June 2006 and includes among other things the latter text, the parts of the 1988 Accord that were not revised and the 1996 Amendment for market risk.[5]

The current Basel framework consists of four parts. The structure is depicted in **FIGURE 1**. In the first part, the scope of the rules is discussed and some general remarks are made. The next three parts are referred to as the three pillars. The first pillar deals with the calculation of minimum capital requirements for three broad risk categories: credit risk, operational risk and market risk. The second pillar contains rules with regard to the supervisory review. In paragraph 720 of the comprehensive document, it is mentioned that the purpose of the supervision is "not only to ensure that banks have adequate capital to support all the risks in their business, but also to encourage banks to develop and use better risk management techniques in monitoring and managing their risks". Pillar 3 is meant to complement the first two pillars. It contains a number of disclosure recommendations and requirements in order to give market participants a sound overview of the risk position of a bank.

[5] When we refer to the Basel framework in the remainder of the text, then the comprehensive version of the Accord is meant, which is the current version at the moment of writing this thesis.

Figure 1 | **Structure of the current Basel framework. Source: Basel Committee on Banking Supervision (2006, p. 6)**



Notes: This figure depicts the structure of the document 'International convergence of capital measurement and capital standards: a revised framework: comprehensive version' graphically. The marked rectangle represents the part of the first pillar – minimum capital requirements – that deals with the measurement of market risk and the calculation of the corresponding capital charge.

The part of the Basel framework that is especially relevant in the context of this thesis, is the section of the first pillar that deals with market risk (marked red in **FIGURE 1**).[6] Two methods are proposed to measure market risk. The first alternative is a standardized method, which is set out in the framework itself. The second alternative makes it possible for a bank to use its own internal Value-at-Risk model to measure market risk. This method ('the internal models approach') is only allowed after the explicit approval of the supervisor; a certain number of qualitative and quantitative criteria, summed up in paragraph 701(ii) of the comprehensive document, should be fulfilled. One of the qualitative criteria to be met by a bank comprises the establishment of a risk control unit that regularly examines the quality of the internal model. The quality is judged by comparing the daily Value-at-Risk estimate with the trading result of the subsequent day. This ongoing process is referred to as backtesting and will be the main subject in the empirical part of this thesis. A more in-depth description of the backtesting procedure is given in Annex 10a of the comprehensive document.

[6] This subject is mainly covered in paragraph 683(i) up to and including paragraph 718(xcic).

Table 1 | **Penalty scheme**

Zone	Number of exceptions	k	Cumulative probability
Green	0	3,00	8,11%
	1	3,00	28,58%
	2	3,00	54,32%
	3	3,00	75,81%
	4	3,00	89,22%
Yellow	5	3,40	95,88%
	6	3,50	98,63%
	7	3,65	99,60%
	8	3,75	99,89%
	9	3,85	99,97%
Red	10 or more	4,00	99,99%

Notes: In this table the penalty zones as identified by the Basel Committee on Banking Supervision are shown. The backtesting procedure is carried out based on a sample of the observations of the 250 preceding days. According to the number of exceptions in this time frame the backtesting result is classified in three zones. These zones are depicted in the first column. The green zone implies that the number of exceptions is acceptable and there is no reason for an adjustment. When the number of exceptions is in the yellow zone, then the result is ambiguous. It might indicate that the model is wrongly specified or it could be more coincidence. The second column shows the actual number of exceptions and the third column gives the corresponding hysteria factor. The fourth column depicts the probability that a given number of exceptions or a smaller number occurs (with a true coverage of 99%).

After the measurement of the market risk a capital charge is calculated. The capital charge depends on the type of model that is applied by a bank. In case of the standardized approach the capital requirement is equal to the sum of (1) a charge based on specific risks of securities and (2) a charge based on the interest rate risk. In case of the internal models approach the capital charge is equal to the higher of (1) the VaR of the previous business day and (2) the average of the VaR over the last sixty business days, multiplied by a factor k. In formula, the capital charge based on the latter model can be described as follows:

$$MRC_t = Max\left[k\,\frac{1}{60}\sum_{i=1}^{60} VaR_{t-i}^{c}\ , VaR_{t-1}^{c}\right] \tag{1}$$

where MRC_t is the market risk charge on day t and c is the confidence level, which is set at 99% by the Basel Committee. The multiplication factor k is also called ‘hysteria factor’. The hysteria factor depends on the outcome of the backtesting procedure. According to the number of exceptions in the past 250 days the backtesting result is classified in a zone. Three zones are defined; these are depicted in the first column of **TABLE 1**. The third column shows the hysteria factor associated with a certain number of exceptions. As can be seen, the minimum multiplication factor is 3. The specific classification is based on statistical considerations. Indeed, it is not possible to judge the quality of a model based on the number of exceptions with complete certainty. Inference could be clouded by statistical errors.

The total capital requirement of a bank is equal to the sum of the aforementioned market risk charge, a charge for operational risk and a charge for credit risk. Equivalently, this capital requirement or 'capital base' should be equal to 8% of the total risk-weighted assets. In other words (Basel Committee on Banking Supervision, 2006, p. 12): "Total risk-weighted assets are determined by multiplying the capital requirements for market risk and operational risk by 12.5 (i.e. the reciprocal of the minimum capital ratio of 8%) and adding the resulting figures to the sum of risk-weighted assets for credit risk."

Three types of bank capital contribute to the total capital requirement. First, at least 50% of the capital base should consist of tier 1 capital or 'core capital', being equity and disclosed reserves. Tier 2 and tier 3 capital can supplement the core capital. Tier 2 capital or 'supplementary capital' refers to undisclosed reserves, revaluation reserves, general provisions, hybrid debt capital instruments and subordinated term debt. Tier 3 capital consists of short-term subordinated debt and can only be applied to cover the market risk charge arising from equation (1). Thus, banks are not allowed to cover (part of) their credit risk charge with tier 3 capital.

2.2 THE VALUE-AT-RISK DEBATE

The term Value-at-Risk has been mentioned several times already, but we have omitted to explore the statistical concept in depth thus far. Therefore, we now continue with a verbal definition of Value-at-Risk and a discussion of the merits and shortcomings of the measure.

Jorion (2007, p. 17) provides the following definition of Value-at-Risk: "VaR summarizes the worst loss over a target horizon that will not be exceeded with a given level of confidence."

Although Value-at-Risk has been accepted as a risk measurement tool worldwide in recent years, the concept is not undisputed. The controversy around Value-at-Risk has led to a famous debate in the journal Derivatives Strategy with Philippe Jorion, expert on the topic of Value-at-Risk, as proponent, and Nassim Taleb, economist and literary essayist, as opponent of Value-at-Risk.[7] In this debate several aspects of Value-at-Risk were commented by both authors. Jorion (2007) reviews this debate and enumerates some of the 'side effects' of Value-at-Risk.

[7] Jorion is the author of the book 'Value at Risk: the new benchmark for managing financial risk' (2007) and professor of Finance at the University of California. Taleb is the author of the book 'The Black Swan: the impact of the highly improbable' (2007) and professor of Risk Engineering at New York University. The cause of the debate was an interview with Taleb in an earlier issue of Derivatives Strategy. Jorion replied on the statements made in this interview and in turn Taleb responded.

The first side effect he mentions (Jorion, 2007, p. 551-552) bears on the criticism, among others discussed in Jorion and Taleb (1997), that Value-at-Risk is useless because it is not perfect. Proponents of Value-at-Risk admit that the model is not perfect and that it is impossible to calculate an estimate of the absolute worst outcome. Indeed, only a range of outcomes can be provided. However, the proponents object that Value-at-Risk is useless, because it still gives an approximation of the risk to be expected. Opponents claim that there is a internal contradiction in this line of thought. They state that it is impossible to control risks with a measure that is risky itself and only works on average.

Another side effect brought forward by Jorion (2007, p. 552) concerns "the deliberate creation of risky trades that appear to be low risk in a VaR framework", which he calls VaR arbitrage. This problem might occur when traders are subject to a certain VaR limit. According to this hypothesis, traders try to avoid the VaR limit in order to maximize returns.

A third side effect (Jorion, 2007, p. 554-556) might be especially relevant in the context of this thesis. It bears on the hypothesis that systemic risk is increased by the widespread application of VaR models. This hypothesis is based on the following line of though. If volatility increases as a result of the course of events, then the banks' capital requirements demanded by the supervisor (based on the Basel regulation) also increase. Banks can choose to meet these requirements or to unwind positions. If they choose the latter, then prices will fall and volatility increases further. This will set in motion a vicious circle, which might end with a systemic crisis.

Finally, proponents add that Value-at-Risk should be accompanied by stress tests to take into account extreme events. This is also propagated by the Basel Committee on Banking Supervision (2006). However, the quality of these stress tests has been the subject of attention during the financial crisis. In short we can conclude that proponents of the Value-at-Risk approach argue that potential shortcomings of the model should be fixed by adding other methods, such as stress tests. Opponents, on the other hand, reject this notion and are of the opinion that the model as it is now should be banned. Our contribution to this discussion, in the light of our own empirical findings, will be made in the concluding remarks.

2.3 EMPIRICAL RESEARCH

Although Value-at-Risk as a risk measure has been the subject of a substantial amount of academic papers, the number of empirical studies of internal models used by banks is quite limited because little information concerning these models is disclosed (Berkowitz and O'Brien, 2002; Pérignon, Deng and Wang, 2008). Besides this type of studies, there are several other papers that provide evidence on the quality of different VaR specifications by applying them to historical data (e.g., Hendricks, 1996). Furthermore, some more theoretical papers investigate the moral hazard problem accompanying the internal models approach.

Among others, Lucas (2001) addresses the latter problem. As mentioned, according to the Basel framework banks are allowed to apply their own internal models to calculate the capital requirement for market risk. The objective of the supervisor is to ensure that the banks' internal models are accurate. For this purpose, they have created the backtesting procedure we just discussed.[8] However, for banks it may be beneficial to apply a model that is not accurate (i.e., a model that is more or less conservative than it should be). These opposed interests between supervisor and banks create a moral hazard problem. Lucas (2001) investigates whether or not banks have an incentive in the current regulatory framework to use a model that is not correctly specified. He finds that "the monetary penalties are too low to provide a sufficiently strong incentive for banks to design internal risk management models that produce good estimates of their true VaR" (Lucas, 2001, p. 842), implicating that banks are provided with an incentive to understate their market risk. According to Lucas (2001), banks can be induced to increase the coverage of their internal models by adjusting the penalty scheme. He suggests to increase the penalties and to make the penalty scheme "steeper".

The first direct evidence on the accuracy of internal risk models used by commercial banks was published by Berkowitz and O'Brien (2002). They gathered data on the daily trading revenue and the associated daily 99th percent Value-at-Risk of six anonymous commercial banks in the United States. The descriptive statistics (Berkowitz and O'Brien, 2002, p. 1096) already reveal that the number of exceptions is lower than expected, which indicates that the VaR estimates of the banks are too conservative. However, when a exception occurs then its magnitude is larger than what would be expected based on the normal distribution.

Berkowitz and O'Brien (2002) also provide more formal tests of the quality of banks' internal models.

First, they apply two different coverage tests to examine the accuracy of the banks' internal models. These tests will be discussed more extensively in chapter 3. For now, we will confine ourselves with some general remarks. The unconditional coverage test compares the

[8] See **Table 1** for a detailed overview of the penalty scheme.

exception rate that should occur based on the model specification with the actual exception rate. The results (Berkowitz and O'Brien, 2002, p. 1102) show that every bank has an exception rate lower than 1%. One of these banks has an exception rate that is significantly lower than 1% based on the unconditional coverage test. For another bank, the unconditional coverage test is undefined because this bank has no exceptions at all. Because of the shortcomings of the unconditional coverage test, another test is carried out. This test, the conditional coverage test, not only compares the theoretical exception rate with the actual exception rate, but also takes into account the distribution of the exceptions over time. Based on the unconditional coverage test, the internal models of two banks are rejected. Again, for one bank the test is undefined because the exception rate is zero.

Second, Berkowitz and O'Brien (2002) compare the banks' VaR figures with estimates produced by a reduced form model. This benchmark consists of an ARMA(1,1) and a standard GARCH(1,1) model and is based on the time series of profits and losses (arising from the trading book). Again, unconditional and conditional coverage tests are performed, but now for the benchmark model. The results of these backtests (Berkowitz and O'Brien, 2002, p. 1105) indicate that in the present case a time series model performs just as well as the banks' internal models. Both the unconditional and the conditional coverage test point out that the time series model should be rejected for one (and the same) bank. Although the performance across the different models is fairly similar in terms of coverage, the authors also find that the ARMA(1,1)-GARCH(1,1) is better able to capture substantial changes in volatility of the profits and losses.

After presenting the empirical results, Berkowitz and O'Brien (2002) discuss some potential causes for the conservativeness of the banks' internal models.

First, they find that banks include net fee income in the reported profits and losses arising from their trading activities, but exclude it from their Value-at-Risk estimates. The average effect of this policy is unknown because net fee income is not disclosed by banks. However, if the profit and loss series would be corrected for this bias, then the exception rate would come closer to the targeted rate.[9]

Second, the risk overstatement could be caused by neglecting possible diversification benefits in the process of calculating Value-at-Risk for a portfolio as a whole. The aggregate VaR

[9] This subject is also discussed by Hendricks and Hirtle (1997, p. 6). They state that the inclusion of fee income from intraday trading in the trading result has two opposite effects. On one hand, the number of exceptions decreases because the trading result increases. On the other hand, the number of exceptions might increase because intraday trading leads to a higher volatility of the trading result.

of a bank is often calculated as the sum of several 'subgroup' VaRs. The term 'subgroup' then refers to a business line or risk class.[10] According to Berkowitz and O'Brien (2002, p. 1109), "of the banks whose VaRs are among the most conservative, several make extensive use of the subportfolio addition procedure". However, a quick look at some recent annual reports reveals that most large banks nowadays take into account diversification effects. For example, Deutsche Bank (2007, p. 87) states: "Diversification effect reflects the fact that the total Value-at-Risk on a given day will be lower than the sum of the Values-at-Risk relating to the individual risk classes. Simply adding the Value-at-Risk figures of the individual risk classes to arrive at an aggregate Value-at-Risk would imply the assumption that the losses in all risk categories occur simultaneously."

Third, Berkowitz and O'Brien (2002) argue that banks' internal models might be conservative because of regulatory requirements. An interesting aspect concerns the supervisory backtesting procedure. As mentioned earlier in the section regarding the Basel framework, the backtesting result (in the case of the internal models approach) is classified in a zone according to the number of exceptions in the past 250 days. If this number is in the yellow or red zone, the capital charge increases with a certain amount. This practice may provide banks with an incentive to overstate their risks because by doing so an exception is less likely to occur.

Berkowitz and O'Brien (2002) conclude that the internal market risk models of banks are (too) conservative. Pérignon, Deng and Wang (2008) present a more quantitative method to measure the exact amount of risk overstatement. Their empirical analysis is based on a sample of six Canadian commercial banks, which publish graphs of the daily VaR and trading revenue in their annual reports. Pérignon *et al.* (2008) convert these graphs into two time series (one for the daily VaR and one for the trading revenue) using a so called data extraction method. This method makes it possible to explore the quality of banks' internal models while keeping the identity of these banks intact (contrary to the anonymous data of Berkowitz and O'Brien, 2002). The descriptive statistics (Pérignon *et al.*, 2008, p. 788) show that the number of exceptions is much lower than expected, which indicates that the VaR estimates of the banks are too conservative. This is in line with Berkowitz and O'Brien (2002), although the understatement in Pérignon *et al.* (2008) looks more severe at first sight.

[10] Regarding business lines, Berkowitz, Christoffersen and Pelletier (2008) show for a large commercial bank that it calculates Value-at-Risk separately for each of its desks. Regarding risk categories, interest rate risk, foreign exchange risk, equity risk and commodity risk are among the most commonly used risk classes.

Pérignon *et al.* (2008) also perform formal tests to judge the quality of banks' internal models.

In the first place, three statistical tests - two unconditional coverage tests and one conditional coverage test - are applied to the internal models of the six banks. Every internal model is rejected based on each of these tests, implying "overwhelming" evidence for risk overstatement (Pérignon *et al.*, 2008, p. 790).

Next, the banks' internal models are compared with two benchmark models, one being a historical simulation model and the other a standard GARCH(1,1) model. These models are based on the time series of profits and losses arising from the trading book. As in Berkowitz and O'Brien (2002), the benchmark models are less conservative than the bank models. Both the historical simulation model and the GARCH(1,1) model produce a number of exceptions that approaches the targeted number. The three coverage tests just mentioned are also applied to the benchmark models. Both models are acceptable (for every bank) based on each of these tests (Pérignon *et al.*, 2008, p. 790).

After providing evidence on the conservativeness by banks, Pérignon *et al.* (2008) continue with quantification of the risk overstatement. First, they calculate a risk overstatement factor ρ and find that "the risk overstatement estimates go from a moderate 19% for National Bank of Canada to an extremely high 79% for Bank of Montreal and Royal Bank of Canada" (Pérignon *et al.*, 2008, p. 791). Subsequently, they estimate the costs associated with this degree of risk overstatement. Indeed, from the banks´ point of view opportunity costs are involved with the superfluous regulatory capital held by banks. These opportunity costs range from 0,4 to 8,3 million Canadian dollars per year (Pérignon *et al.*, 2008, p. 792). Moreover, some other potential costs of risk overstatement are mentioned. For instance, it is argued that an overstated capital charge might distort stock prices and consequently lead to improper portfolio choices. Furthermore, the credit policy of a bank might be influenced, which has a negative effect on the economy as a whole.

Just as Berkowitz and O'Brien (2002) did, Pérignon *et al.* (2008) dedicate a paragraph to a discussion of potential causes for risk overstatement. Besides the ones already mentioned in this thesis, they raise two more causes. It is put forward that banks are conservative on purpose, because an internal model that yields too much exceptions might harm their reputation. Furthermore, discussions with practitioners in the field reveal that risk managers may intentionally make conservative estimates of Value-at-Risk to avoid unwanted attention.

Pérignon and Smith (forthcoming) extend the research of Berkowitz and O'Brien (2002) and Pérignon *et al.* (2008). A part of their empirical analysis is based on a sample of five international commercial banks. They use the same data extraction method as Pérignon *et al.* (2008) in order to convert graphs in annual reports into daily VaRs and profits and losses. Their results confirm the earlier findings that banks' Value-at-Risk estimates are too conservative. All banks show less exceptions than expected. Based on an unconditional coverage tests four out of five internal models are rejected (Pérignon and Smith, forthcoming).

In a number of other studies the performance of Value-at-Risk models is evaluated using historical data. For instance, Hendricks (1996) tests different VaR specifications by applying them to a large number of foreign exchange portfolios during a period of twelve years (daily data). Among the models under investigation are five equally weighted moving average models, three exponentially weighted moving average models and four historical simulation models. The author finds that the 99^{th} percent Value-at-Risk, which is the most commonly used measure in the banking industry, is too low in most cases. "Only the 1250-day historical simulation approach attains 99% coverage across all 1000 portfolios (…). The other approaches cover between 98,2 and 98,8% of the outcomes on average across portfolios" (Hendricks, 1996, pp. 49-50). Although this study follows a different approach, the results contradict those of Berkowitz and O'Brien (2002) and Pérignon *et al.* (2008), who find that the banks' internal models provide too much coverage instead of too little.

Based on the discussion above, we can draw some conclusions. First, the penalty scheme established by the Basel Committee may provide banks with an incentive to understate their true market risk. Second, empirical evidence on the accuracy of banks' internal models reveals that in practice the opposite is true; banks consistently overstate their market risks. This in turn is in contrast with a more general study of the accuracy of 99^{th} percent Value-at-Risk models, which show that most model specifications provide too little coverage.

CHAPTER THREE

DATA ACQUISITION AND METHODOLOGY

In this chapter we will discuss the data acquisition process and methodology applied in our empirical research. First, we explain which banks are part of our sample and how we retrieved the profit and loss figures and Value-at-Risk estimates of these banks. The merits and shortcomings of the data extraction method as developed by Pérignon *et al.* (2008) are reviewed. Second, attention is paid to backtesting procedures. Statistical aspects of different tests are covered. Third, the different benchmark models are defined. And finally, we explain how we measure the degree of risk overstatement and the associated costs.

3.1 THE DATA ACQUISITION PROCESS

In principle, data concerning daily trading revenue and the associated VaR estimates is not disclosed by banks. However, some banks depict this information graphically in their annual reports. Pérignon *et al.* (2008) present a data extraction method[11] to translate graphs into two daily time series (one for the trading revenue and one for the associated VaR estimate). As mentioned earlier, this method makes it possible to examine the quality of banks' internal models while keeping the identity of these banks intact. Pérignon *et al.* (2008) also examine the accuracy of their method by means of simulations. They compare actual data with estimates retrieved through the data extraction method and find a mean absolute error of 0,91% (Pérignon *et al.*, 2008, p. 785).

[11] This data extraction method is based on the application MATLAB. A detailed description of the procedure and an examination of its accuracy are presented in Pérignon *et al.* (2008, p. 793-794) and Pérignon and Smith (forthcoming).

Although the data extraction method provides us with numerical, non-anonymous information regarding trading revenues and VaRs of banks, it also has several limitations.

For backtesting purposes Value-at-Risk estimates should be compared with the hypothetical buy-and-hold profits or losses. However, graphs in banks' annual reports are often based on actual profits and losses of the day subsequent to the day the VaR figure is forecast. Furthermore, despite the small mean absolute error reported by Pérignon *et al.* (2008), we still deal with estimates based on graphs. If the graphs themselves are smoothed by banks, then this will directly affect the quality of our data set. Nevertheless, the extent of this potential bias is unknown. Finally, the inclusion of fee income in the graphs could distort the backtesting process (see also footnote 8 on page 11).

In this thesis a sample of five large commercial banks is used to analyze the quality of banks' Value-at-Risk models. A detailed analysis of the time series of these banks, retrieved through the data extraction method, is covered in the next chapter.

3.2 BACKTESTING PROCEDURES

Several statistical tests have been developed to judge the quality of Value-at-Risk models. Some of the most commonly used tests are the unconditional and conditional coverage tests based on a log-likelihood ratio.[12]

The *unconditional* coverage test can be defined as follows (Jorion, 2007, p. 147):

$$LR_{uc} = -2ln[(1-p)^{T-N}p^N] + 2ln\left\{\left[1-\left(\frac{N}{T}\right)\right]^{T-N}\left(\frac{N}{T}\right)^N\right\} \quad (2)$$

with p as the VaR confidence level, T as the number of observations and N as the number of exceptions. The log-likelihood ratio follows a χ^2 distribution with one degree of freedom.[13] Under the null hypothesis the VaR confidence level is equal to the actual exception rate. In other words, if the null hypothesis is rejected, then the risk model is incorrectly specified.

[12] See Christoffersen (2003), Campbell (2005) or Jorion (2007) for an in-depth explanation of this subject.

[13] This is approximately true if the number of observations, T, is sufficiently large. When the number of observations is small, it might be better to use P-values derived from Monte Carlo simulations. Because the time series in this thesis consist of a relatively large number of observations, we rely on the χ^2 distribution. See Christoffersen (2003, p. 184) for details.

Table 2 | **Transition probability matrix**

		Today	
		No exception	Exception
Previous day	No exception	$\pi_{00} = \frac{T_{00}}{T_{00} + T_{01}}$	$\pi_{01}(= \pi_0) = \frac{T_{01}}{T_{00} + T_{01}}$
	Exception	$\pi_{10} = \frac{T_{10}}{T_{10} + T_{11}}$	$\pi_{11}(= \pi_1) = \frac{T_{11}}{T_{10} + T_{11}}$

Notes: In this table the transaction probability matrix is depicted. A day *without* an exception is defined as 0 and a day *with* an exception is defined as 1. T_{ij} $(i, j = 0{,}1)$ is the number of observations with *j* following i and π_{ij} $(i, j = 0{,}1)$ is the probability of *j* on a certain day given i on the previous day. For the sake of completeness, it is noted that π_i is equivalent to π_{ij}, with *j* being 1.

Although the unconditional coverage test is useful for testing the coverage *on average*, it does not take into account the fact that exceptions are often clustered in time. Christoffersen (2003) provides a test statistic to measure whether or not the observed exceptions are serially independent. If we define a day *without* an exception as 0 and a day *with* an exception as 1, then we can write the independence test as follows (Jorion, 2007, p. 151):

$$LR_{ind} = -2ln\left[(1-\pi)^{(T_{00}+T_{10})}\pi^{(T_{01}+T_{11})}\right] + 2ln[(1-\pi_0)^{T_{00}}{\pi_0}^{T_{01}}(1-\pi_1)^{T_{10}}{\pi_1}^{T_{11}}] \qquad (3)$$

where T_{ij} $(i, j = 0{,}1)$ is the number of observations with *j* following i and π_i $(i = 0{,}1)$[14] is the probability of an exception on a certain day given i on the previous day. For example, T_{01} gives the number of times where a day with an exception follows a day without an exception and π_1 is the probability of an exception given an exception the previous day. A complete overview of the possibilities, commonly called transition probability matrix, is displayed in **TABLE 2**. Again, the test statistic follows a χ^2 distribution with one degree of freedom.[15] The null hypothesis is rejected if the exceptions are clustered in time.

The *conditional* coverage test is essentially an extension of the *unconditional* coverage test. It tests whether the risk model is correctly specified on average *and* whether the exceptions are independent in time. The following test statistic corresponds to this test (Jorion, 2007, p. 152):

$$LR_{cc} = LR_{uc} + LR_{ind} = (2) + (3) \qquad (4)$$

[14] For the sake of completeness, it should be noted that π_i is equivalent to π_{ij}, with *j* being 1.

[15] The same reservation is in effect as in the case of the unconditional coverage test. See also footnote 12.

This test statistic follows a χ^2 distribution with two degrees of freedom.[16] The null hypothesis is rejected if either the model is incorrectly specified on average or the exceptions are clustered in time.

3.3 SPECIFICATION OF BENCHMARK MODELS

A number of benchmark models based on the profit and loss series of the sample banks is compared with the banks' internal models. These benchmark models include two models based on historical simulation, an ARMA(1,1)-GARCH(1,1) model and the RiskMetrics model.

The historical simulation approach uses a specified number of historical observations to calculate the n^{th} percentile Value-at-Risk measure. For example, when the historical simulation is based on 250 daily trading results, then the 99% Value-at-Risk is the third largest loss.[17]

The number of days included in the observation period is the result of a tradeoff. Short-term movements in the underlying portfolio are best captured with a short observation period. However, a longer observation period produces more accurate forecasts. In the empirical part of this thesis we apply two simplified historical simulation models, one based on 125 historical trading results and one based on 250 historical trading results. In each case, we determine the 99^{th} percentile risk measure.

As we will see later, the banks' distribution of profits and losses might be characterized by excess kurtosis, or 'fat tails', which could indicate time variation in risk. Thus, in order to capture volatility clustering, we also estimate a simple ARMA(1,1)-GARCH(1,1) model based on the time series of profits and losses. This approach, which is based on Berkowitz and O'Brien (2002, p. 1103), can be described as follows:

$$r_t = \mu + \gamma r_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \tag{5}$$

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{6}$$

$$VaR_t^{99\%} = \hat{r}_t - 2{,}33\sqrt{\widehat{\sigma_t^2}} \tag{7}$$

---

[16] The same reservation is in effect as in the case of the unconditional coverage test. See also footnote 12.

[17] 1% of the sample that is larger (in absolute terms) than the Value-at-Risk measure is equal to 2,5 observations, but since this observation does not exist, the figure is rounded to 3. Theoretically, the third observation is not necessary a loss because the measure is based on the trading result (i.e. profits and losses).

Equation (5) reflects the ARMA(1,1) process (Box and Jenkins, 1976) , where r_t is the profit or loss on day t, μ is the expectation of r_t and the error terms ε_t and ε_{t-1} are standard normal random variables. The GARCH (1,1) process (Bollerslev, 1986) is depicted in equation (6). The parameter ω can be defined as $(1-\alpha-\beta)\sigma^2$, where σ^2 is the unconditional long-run variance. The persistence of the model, which indicates how long it takes before volatility returns to the long-run average, is equal to $\alpha+\beta$. We set a rolling window of 250 business days to estimate the parameters μ, γ and θ of the ARMA(1,1) equation and ω, α and β of the GARCH(1,1) equation. Thus, during the first 250 days one-day ahead forecasts are calculated in-sample. After this period, the parameters are re-estimated daily based on the 250 preceding observations. One-day ahead forecasts are then calculated out-of-sample. Finally, using the estimates $\widehat{r_t}$ and $\widehat{\sigma_t^2}$ and equation (7) the 99% Value-at-Risk measure on day t can be computed. We use the Gaussian distribution to calculate the 99% Value-at-Risk, because it is easy to implement. However, we are aware of the limitations of this approach. If the distribution of r_t turns out to be flatter ('fat tails') or tighter than the Gaussian distribution, then this method will not yield accurate Value-at-Risk estimates (underestimation and overestimation, respectively).

The RiskMetrics model (J.P. Morgan, 1996) can be considered as a restricted GARCH(1,1) model.[18] We obtain the RiskMetrics model from equation (6) if $\alpha=1-\lambda$, $\beta=\lambda$, $\alpha+\beta=1$ and $\omega=0$:[19]

$$\sigma_t^2=\lambda\sigma_{t-1}^2+(1-\lambda)r_{t-1}^2 \tag{8}$$

where σ_t^2 is an exponentially weighted average of the exponentially weighted average of the previous day, σ_{t-1}^2, and the squared return of the previous day, r_{t-1}^2. The parameter λ, which is also called 'decay factor', is the only parameter that has to be estimated. However, because the estimation process requires a lot of time in practice, RiskMetrics uses a fixed decay factor of 0,94 for time series with daily data (Jorion, 2007, p. 232). Since the two time series in this thesis also consist of daily data, we follow this approach and set $\lambda=0{,}94$.

[18] Some authors refer to this model as integrated GARCH(1,1) model or IGARCH(1,1).

[19] Thus, the persistence of the RiskMetrics model equals 1, implying that shocks in volatility persist forever. Long-run variance is not defined in the model.

3.4 MEASURING DEGREE AND COSTS OF RISK OVERSTATEMENT[20]

In section 2.2 we presented a verbal definition of Value-at-Risk. This definition contains two important components, the confidence level and the target horizon over which market risk is measured. Thus, Value-at-Risk is the smallest loss subject to the following condition:

$$P(r_{t+1} < -VaR_t^c) = 1 - c \tag{9}$$

where P is the probability that the trading result on day $t+1$ is smaller than the Value-at-Risk measure forecasted on day t. This probability should be equal to $1-c$, where c is the specified confidence level.

In order to measure the degree of risk overstatement, Pérignon *et al.* (2008, p. 791) introduce the risk overstatement coefficient ρ, which can take on values between 0 and 1. This approach is based on the expectation that banks *overstate* their reported risks. However, strictly speaking it is possible that a bank *understates* its risk. Based on our expectations, we replace the risk *overstatement* coefficient ρ with a risk *deviation* coefficient φ, which can take on *both* positive and negative values. Thus, this coefficient corrects the Value-at-Risk disclosed by a bank ($DVaR_t^c$) such that it is equal to its proper Value-at-Risk (VaR_t^c):

$$VaR_t^c = DVaR_t^c(1-\varphi) \tag{10}$$

If the disclosed Value-at-Risk is larger (smaller) than the true Value-at-Risk (in absolute terms), then the probability of an exception is smaller (larger) than it should be. The risk deviation coefficient adjusts the disclosed Value-at-Risk until the desired confidence level c is reached. We can depict this formally by combining equations (9) and (10):

$$P[r_{t+1} < DVaR_t^c(1-\varphi)] = 1 - c \tag{11}$$

Pérignon *et al.* (2008, p. 791) define their risk overstatement coefficient as "the minimum ρ such that the adjusted time series $DVaR_t^c(1-\rho)$ is not rejected by the unconditional coverage test at a given confidence level". Instead of a single number, we provide three characteristic values of the risk deviation coefficient φ for every bank. Upper and lower limits of the non-rejection range are calculated using equation (2) and a confidence level of 10%. We define φ_{min} as the upper limit in case of risk understatement or as the lower limit in case of risk overstatement. The

[20] The framework for quantifying the degree and costs of model conservativeness is mainly derived from Pérignon *et al.* (2008, p. 791). Nevertheless, it should be noticed that some minor changes in notation are made here.

opposite is true for φ_{max}. Consequently, in either case $|\varphi_{min}|$ is the smallest value of φ in order to reach the non-rejection zone. Furthermore, an exception rate of exactly 1% implies a perfect model. Hence, we also calculate an 'expected' risk deviation coefficient (φ_{exp}) in case of a perfectly specified model. All values can be obtained by performing a large number of unconditional coverage tests for every bank, using increments of $\varphi = 0{,}001$.

As mentioned in section 2.3, from the banks´ point of view costs are involved with the superfluous regulatory capital held by banks. Holding too little regulatory capital is possible as well, but only to a certain extent. Indeed, banks are penalized if their models consistently produce too much exceptions. The market risk charge based on an incorrectly specified model can be quantified using the calculated risk deviation coefficient and the equations (1) and (10). Recall that equation (1) describes how the daily capital requirement is determined. If we substitute $DVaR_t^c(1-\varphi)$ for VaR_t^c using equation (10), then we obtain the following equation:

$$MRC_t^{RD} = Max\left[k\,\frac{1}{60}\sum_{i=1}^{60}\frac{VaR_{t-i}^c}{1-\varphi}\,,\frac{VaR_{t-1}^c}{1-\varphi}\right] \tag{12}$$

where MRC_t^{RD} is the market risk charge on day t given a too conservative or too tight model, k is still the hysteria factor and c the confidence level. Thus, there exists a direct relationship between the market risk charge based on a correctly specified model and the market risk charge based on a model which is too conservative or too tight. This relationship, which depends solely on the risk deviation coefficient φ, can be written as follows:

$$MRC_t^{RD} = \frac{1}{1-\varphi}MRC_t \tag{13}$$

Recall that φ_{min} is the smallest value (in absolute terms) of φ in order to reach the non-rejection zone. If we insert this value and the value of MRC_t^{RD} in equation (13) for each bank, we can calculate MRC_t. Subsequently, it is possible to calculate the excess market risk charge:

$$Excess\ MRC_t = MRC_t^{RD} - MRC_t \tag{14}$$

Strictly speaking the 'excess' market risk charge can take on a negative value. This happens when a model is too tight. However, when calculating the monetary effects of an incorrectly specified model we only consider costs involved and thus positive values of the excess market risk charge. Potential 'benefits' of a model with too much exceptions are left out of consideration because eventually the supervisor will not allow the occurrence of such a situation.

If we multiply the excess market risk charge by the opportunity costs of capital that are in force for the different banks, we obtain the monetary costs of risk overstatement. We follow Pérignon *et al.* (2008) and use the return on assets (ROA) as a proxy for the opportunity costs of capital. This results in the following equation:

$$MC = (Excess\ MRC) \times ROA \tag{15}$$

where MC represents the annual monetary costs.

Now we have outlined the source of the data and the methodological framework, we are able to continue with a thorough discussion of the data set and the results of our empirical research. This will be the subject of the next chapter.

CHAPTER FOUR

DATA ANALYSIS AND DISCUSSION OF THE RESULTS

This chapter contains an in-depth analysis of the time series derived from the graphs of the annual reports of five banks. Furthermore, the quality of the banks' risk models will be the subject of investigation. The validity of these models will be judged using several statistical tests. The risk deviation factor is applied to measure the degree of risk deviation.

4.1 DATA ANALYSIS

The empirical part of this thesis is based on a sample of five large commercial banks, being ABN AMRO (the Netherlands), BNP Paribas (France), Credit Suisse (Switzerland), Deutsche Bank (Germany) and Nordea (Sweden). Each of the banks is established in Europe and internationally active. Time series of trading revenues and VaR estimates are obtained by means of the data extraction method for the years 2002 up to and including 2008.[21] To test whether the original graphs coincide with the graphs based on the extracted data, we place the latter graphs upon the former. If any deviations between the graphs are still present, we correct the underlying time series.

[21] Not every bank in the sample publishes the necessary graphs throughout the whole sample period. ABN AMRO and Credit Suisse provide graphs in the annual reports of 2002, 2003, 2004 and 2005. The sample period of BNP Paribas is the same, but no graph is published in the annual report of 2003. Nordea presents graphs in 2004, 2005, 2006 and 2007. Deutsche Bank publishes graphs during the entire sample period. Furthermore, on a few days in the sample it was not completely clear whether an exception occurred because the trading revenue and VaR intersected in the graph. If such a case of doubt occurred we did *not* count it as an exception.

Table 3 | **Summary statistics sample banks**

	ABN	BNP	CS	DB	NOR[e]
		Panel A: Description of VaR models, as of Dec. 2008			
Measurement method	Historical simulation	GEaR (Gross Earnings at Risk)	Historical simulation	Monte Carlo simulation	Historical simulation
Moving window	1,5 years	260 trading days	2 years	261 trading days	2 years
Confidence level	99%	99%	99%	99%	99%
Backtesting method	Actual and hypothetical P&L	Actual P&L	Actual P&L	Hypothetical P&L	Simulated P&L
		Panel B: Description of key figures, as of Dec. 2008			
Book value of assets (in millions)	EUR 666.817	EUR 2.075.551	CHF 1.170.350	EUR 2.202.423	EUR 474.074
Book value of equity (in millions)	EUR 17.123	EUR 58.968	CHF 32.302	EUR 31.914	EUR 17.803
Book value of derivatives notional (in millions)	EUR 5.260.535	EUR 36.349.436	CHF 40.416.000	EUR 50.432.203	EUR 3.770.087
Return on equity (in %)	Not reported	6,6%	-21,1%	-11,1%	15,3%
		Panel C: Description of capital adequacy ratios and figures, as of Dec. 2008			
Total capital base (in millions)	EUR 25.405	EUR 59.499	CHF 46.090	EUR 37.396	EUR 20.326
Total risk-weighted assets (in millions)	EUR 176.028	EUR 527.643	CHF 257.467	EUR 307.732	EUR 168.572
of which market risk equivalents (market risk charge divided by 8%)	EUR 13.069	Not reported	CHF 39.911	EUR 23.496	EUR 5.930
Market risk charge (in millions, multiple of VaR)	EUR 1.046	Not reported	CHF 3.193	EUR 1.880	EUR 474
Total capital ratio (in %)	14,4%	11,2%	17,9%	12,2%	12,1%

Notes: This table depicts some decriptive statistics with respect to the banks under investigation (ABN AMRO, BNP Paribas, Credit Suisse, Deutsche Bank and Nordea). The information in the table is based on the banks' annual reports of 2008. Panel A gives a summary of the banks' internal Value-at-Risk models. Some of the banks have modified their model during the sample period. Most changes relate to the length of the moving window. ABN AMRO applies a so called 'exponential decay method', implying that the data points are weighted. In panel B some key figures are shown. It should be noted that 'book value of derivatives notional' refers to the notional amounts held for trading purposes. Only in the case of Deutsche Bank, derivatives held for hedging purposes are also included in the amount. Panel C provides a description of the banks' capital adequacy. 'Total capital base' refers to the banks' actual regulatory capital. It should be at least 8% of the risk-weighted assets. Nordea reports that its capital adequacy figures were based on rules that applied before the transition to Basel II.

Summary statistics with respect to the banks in the sample are depicted in **TABLE 3**. Some remarks are in order.

Regarding the banks' internal models, it should be noted from panel A that most of the banks use historical simulation to measure Value-at-Risk. Only Deutsche Bank reports that it applies Monte Carlo simulation. Furthermore, not every bank employs the necessary hypothetical buy-and-hold profits and losses in the backtesting process. For example, BNP Paribas and Credit Suisse report graphs of the actual profits and losses.

Table 4 | **Summary statistics extracted Value-at-Risk and P&L figures**

	ABN	BNP	CS	DB	NOR
			Panel A: VaR		
Mean	-30,65	-35,31	-62,00	-72,34	-13,90
Median	-26,34	-32,29	-62,59	-67,20	-12,76
Minimum	-78,59	-90,00	-93,77	-173,30	-27,67
Maximum	-10,81	-12,13	-26,57	-29,44	-5,29
Standard deviation	13,37	13,73	12,84	26,33	4,78
Skewness	-1,34	-1,59	0,10	-0,80	-0,60
Kurtosis	4,14	6,05	2,78	3,31	2,60
Jarque-Bera test	**362,96****	**629,50****	3,75	**196,41****	**64,16****
Expected exceptions	10	8	10	18	10
Actual exceptions	4	0	2	28	6
			Panel B: Profits and losses		
Mean	1,31	9,83	4,58	37,35	0,27
Median	0,98	9,10	4,48	38,27	0,02
Minimum	-38,82	-29,98	-104,68	-356,77	-22,42
Maximum	52,30	59,48	95,63	566,51	27,24
Standard deviation	10,05	11,27	18,03	53,40	4,57
Skewness	0,14	0,46	-0,22	-0,87	0,06
Kurtosis	4,45	4,74	6,13	16,95	6,76
Jarque-Bera test	**92,89****	**125,95****	**428,45****	**14582,38****	**575,34****
99^{th} percentile	-23,72	-16,78	-46,84	-148,71	-13,00

Notes: This table depicts descriptive statistics with respect to the extracted time series (Value-at-Risk and trading revenue) of the banks under investigation (ABN AMRO, BNP Paribas, Credit Suisse, Deutsche Bank and Nordea). Panel A gives several characteristics of the banks' Value-at-Risk series and panel B provides information on the profits and losses, or trading revenue, of the banks. The Jarque-Bera test follows χ^2 distribution with two degrees of freedom. The sign ** denotes statistical significance at the 1% level.

Panel B of **TABLE 3** shows some general characteristics of the sample banks. There is quite a variation in the book value of assets, the book value of equity and the derivatives notional across the banks. Especially the differences in the size of the derivatives notional are large. Regarding the return on equity, there is also a striking variation across the different banks. The ratios are negative for Credit Suisse and Deutsche Bank, implying net losses which are most likely caused by the financial crisis persisting in 2008. On the other hand, the return on equity of Nordea is considerably high given the circumstances.

In panel C of **TABLE 3** a description of the capital adequacy figures is presented. In the context of this thesis the market risk charge, which is a multiple of the estimated Value-at-Risk, is especially relevant. Credit Suisse and Deutsche Bank were faced with the highest market risk charge at the time of measurement. This corresponds to their large positions in derivative instruments.

Summary statistics with respect to the extracted time series (Value-at-Risk and trading revenue) are displayed in **TABLE 4**. Plots of these time series are presented in **FIGURE 2** and an overview of the exceptions is shown in **FIGURE 3**.

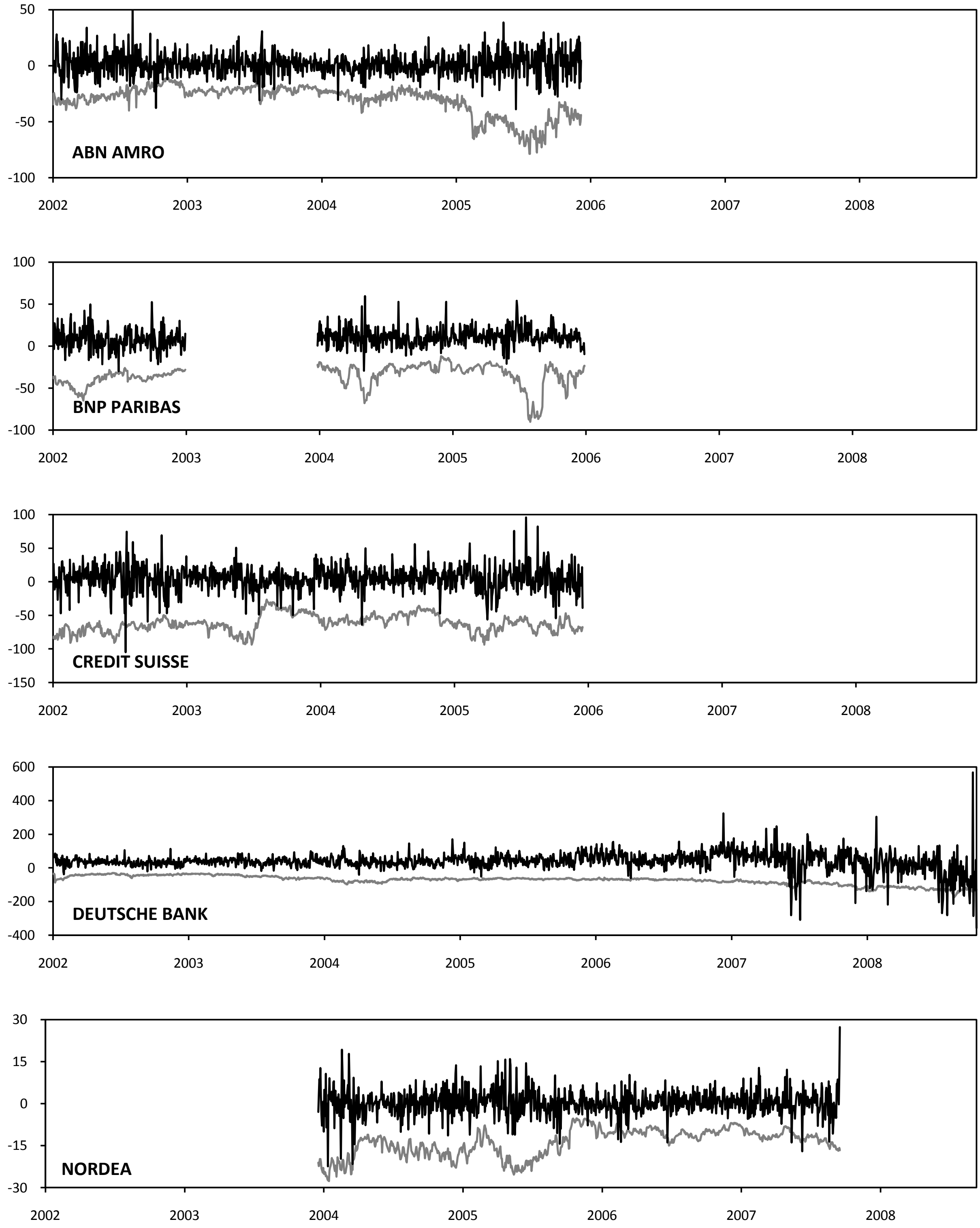
TABLE 4 shows that the mean Value-at-Risk is highest for Credit Suisse and Deutsche Bank, which is in line with our expectation given the previous statistics, and lowest for Nordea. According to the Jarque-Bera test, the Value-at-Risk estimates are only normally distributed in the case of Credit Suisse. Furthermore, a first glance at the number of exceptions already reveals that most of the banks' internal models are rather conservative. For every bank except Deutsche Bank, the actual number of exceptions is lower than the expected number.

The average of the profits and losses of Deutsche Bank is particularly high relative to the other banks, but it is also accompanied by a high standard deviation. The results of the Jarque-Bera test indicate that for none of the banks the profits and losses are normally distributed, which is primarily due to excess kurtosis (especially in the case of Deutsche Bank). The 99th percentile of profits and losses should be comparable to the average of the Value-at-Risk estimates. However, if we compare these figures, we see that the 99th percentile is smaller for every bank.

A look at **FIGURES 2 AND 3** reveals that most of the banks' exceptions occurred in 2007 and 2008, even when there is data available for only two of the five banks in the sample. Especially Deutsche Bank experienced some large losses in 2007 and 2008. According to Deutsche Bank (2007, p. 88) itself "the majority of these outliers were related to extreme events outside standard market conditions", hereby referring to the ongoing international financial crisis. Furthermore, Nordea reports at least four outliers during the year 2006.

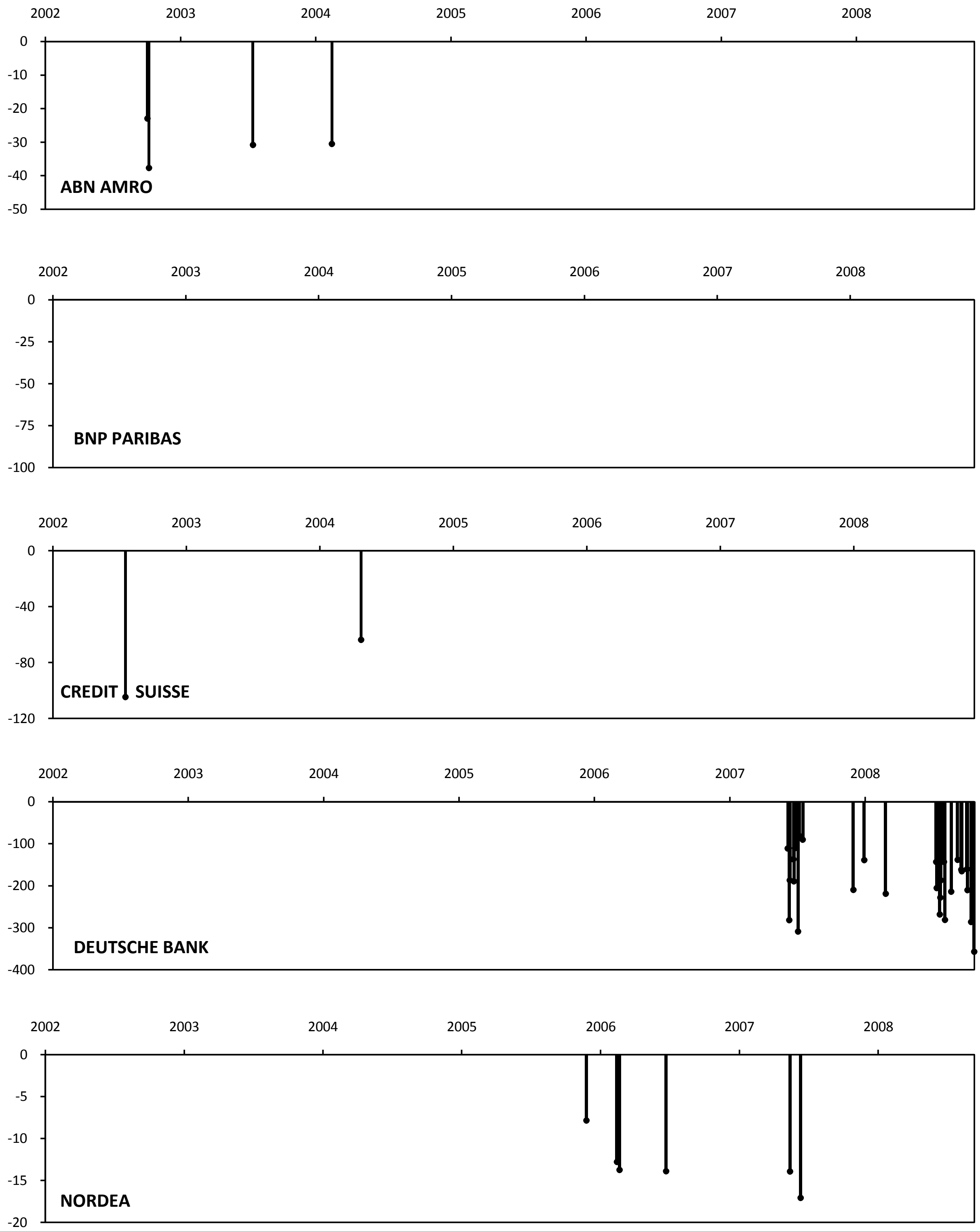
Most likely, these are caused by the market turmoil at that time (mainly in May and June). Thus, most of the exceptions in our data set can be explained by exceptional market movements.

Figure 2 | **Graphical representation of banks' internal VaR models**



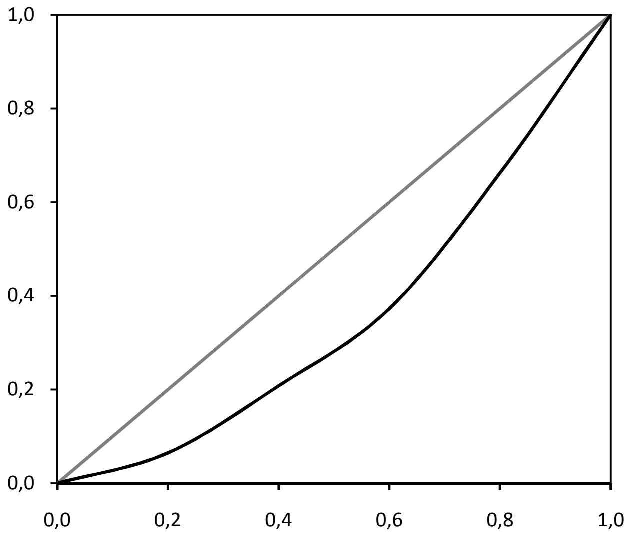
Notes: This figure shows the extracted time series of the banks' daily Value-at-Risk (grey line) and trading revenue (black line). Time (in years) is displayed on the horizontal axis. Daily Value-at-Risk/trading revenue is displayed on the vertical axis, in millions of CHF for Credit Suisse and in millions of EUR for the other banks. The total sample period ranges from January 2002 through December 2008. However, not every bank in the sample publishes graphs necessary to extract data throughout the whole sample period (see also footnote 21).

Figure 3 | **Graphical representation of banks' exceptions**



Notes: This figure shows for every bank the exceptions that occurred during the sample period. Time (in years) is displayed on the horizontal axis. The size of the exception is displayed on the vertical axis, in millions of CHF for Credit Suisse and in millions of EUR for the other banks. The total sample period ranges from January 2002 through December 2008. However, not every bank in the sample publishes graphs necessary to extract data throughout the whole sample period (see also footnote 21).

Figure 4 | **Lorenz curve of the banks' average VaR during the sample period**



Notes: This figure shows the concentration of risks among the five sample banks. The horizontal axis represents the cumulative share of banks from lowest to highest average Value-at-Risk. The vertical axis depicts the cumulative share of aggregated risks. Based on an equal distribution (grey line), it is expected that a random cumulative share of banks is equal to the same cumulative share of aggregated risks. Obviously, this is not the case (black line). For example, the three banks with the smallest average Value-at-Risk (the cumulative share of banks is 60% in that case) do not represent 60% of the aggregated risks but only 37%.

Finally, a Lorenz curve is shown in **FIGURE 4**. This curve displays the concentration of risks among the banks in our sample. Based on an equal distribution (grey line), it is expected that a random cumulative share of banks is equal to the same cumulative share of aggregated risks. Obviously, this is not the case (black line). For example, the three banks with the smallest average Value-at-Risk (thus, a cumulative share of 60%) do not represent 60% of the aggregated risks but only 37%.

4.2 BACKTESTING RESULTS

Now we have reviewed several features of our data set, we are able to continue with a discussion of the quality of the banks' internal models. For that purpose, we also compare the internal models with the benchmark models proposed in the previous chapter.

Statistical information with respect to the accuracy of the banks' internal models and the benchmark models is depicted in **TABLE 5** and **FIGURES 5A AND 5B**.

Table 5 | **Backtesting results banks' internal models and benchmark models**

Bank	Observations	Expected expections	Exceptions based on model	Violation rate	Unconditional coverage test (LR_{UC})	Independence test (LR_{IND})	Conditional coverage test (LR_{CC})
				Panel A: VaR disclosed by banks			
ABN	1022	10	4	0,39%	**4,96***	-	-
BNP	781	8	0	0,00%	-	-	-
CS	1030	10	2	0,19%	**10,10****	-	-
DB	1770	18	28	1,58%	**5,16***	**43,10****	**48,25****
NOR	975	10	6	0,62%	1,68	-	-
				Panel B: 250-day historical simulation			
ABN	772	8	12	1,55%	2,06	-	-
BNP	281	3	4	1,42%	0,47	-	-
CS	780	8	11	1,41%	1,18	**7,20****	**8,39***
DB	1520	15	30	1,97%	**11,36****	**13,92****	**25,28****
NOR	725	7	13	1,79%	3,74	-	-
				Panel C: 125-day historical simulation			
ABN	897	9	20	2,23%	**10,18****	-	-
BNP	531	5	10	1,88%	3,36	1,85	5,21
CS	905	9	17	1,88%	**5,62***	**14,63****	**20,25****
DB	1645	16	37	2,25%	**19,17****	**6,80****	**25,97****
NOR	850	9	17	2,00%	**6,67****	-	-
				Panel D: ARMA(1,1)-GARCH(1,1)			
ABN	1021	10	19	1,86%	**6,11***	0,83	**6,95***
BNP	779	8	8	1,03%	0,01	-	-
CS	1029	10	23	2,24%	**11,76****	**10,59****	**22,35****
DB	1769	18	36	2,04%	**14,75****	**7,67****	**22,42****
NOR	974	10	20	2,05%	**8,39****	-	-
				Panel E: RiskMetrics			
ABN	772	8	11	1,42%	2,06	-	-
BNP	281	3	0	0,00%	-	-	-
CS	780	8	10	1,28%	0,58	-	-
DB	1520	15	8	0,53%	**4,16***	**4,66***	**8,81***
NOR	725	7	12	1,66%	2,64	-	-

Notes: This table shows the backtesting results of the banks' internal models (panel A) as well as the benchmark models (panels B through E). The unconditional coverage test and independence test in the sixth and seventh column both follow a χ^2 distribution with one degree of freedom. The conditional coverage test in the last column, which is the sum of the unconditional coverage test and the independence test, follows a χ^2 distribution with two degrees of freedom. The χ^2 critical values for one and two degrees of freedom are, respectively, 3,841 and 5,991 using a significance level of 5% (denoted by *) or 6,635 and 9,210 using a significance level of 1% (denoted by **). All statistically significant results are in bold type. A dash is shown when the test statistic is not defined.

Panel A of **TABLE 5** reports the results for the banks' internal models. The violation rate of the different proprietary models is depicted in the fifth column. Ideally, this percentage is equal to 1% (given a 99% Value-at-Risk model), but none of the banks' models meets this condition. In case of four models the violation rate is lower than 1%, indicating risk *over*statement. BNP Paribas reports not even a single exception during 781 trading days while we would expect 8 exceptions.

On the other hand, Deutsche Bank reports more exceptions than expected, indicating risk *under*statement. Unconditional coverage tests give rise to a rejection of four of the five models.[22] If we also consider the conditional coverage test, then the model of Deutsche Bank is rejected based on a highly significant test statistic.[23] Indeed, this is caused by the clustering of exceptions in time. Thus, based on the backtesting results only one out of five internal models is accepted.

Panel B and panel C of **TABLE 5** as well as **FIGURE 5A** show the results of the two historical simulation models. It should be noted that both the 250-day and 125-day historical simulation models produce violation rates higher than 1%. In three cases (each regarding the 125-day model) the violation rate is even higher than 2%. Only one out of five (Deutsche Bank) 250-day historical models is rejected based on unconditional coverage tests. If we take exception clustering into account, the model of Credit Suisse is rejected too. Clearly, the 125-day model is less accurate. Based on the unconditional coverage tests only the model of BNP Paribas is accepted. This result is confirmed by the conditional coverage tests. The plots in **FIGURE 5A** graphically show that the time series based on the historical simulation models do not follow the time series based on the banks' internal models. Especially in the case of Deutsche Bank extremely large differences start to exist in 2007 between the outcomes of the historical simulation models and the bank's internal model.

Numerical results of the ARMA(1,1)-GARCH(1,1) model, which is considered the most sophisticated benchmark model, are shown in panel D of **TABLE 5** and plots of the time series (red lines) can be found in **FIGURE 5B**. A detailed overview of the model's parameter estimates is presented in **TABLE 6**. Contrary to our expectations, the ARMA-GARCH model does not yield very accurate Value-at-Risk models. Again, the violation rate is larger than 1% for all banks. The null hypothesis that a bank's model is correctly specified is rejected for four banks based on the unconditional coverage test. Only in the case of BNP Paribas the model produces exactly the expected number of exceptions. However, a test statistic based on the conditional coverage test is not defined for this bank.

[22] Obviously, the model of BNP Paribas is also rejected. However, the test statistic is undefined because no exception is reported.

[23] We cannot judge the quality of the model of Nordea based on the conditional coverage test because the test statistic is undefined.

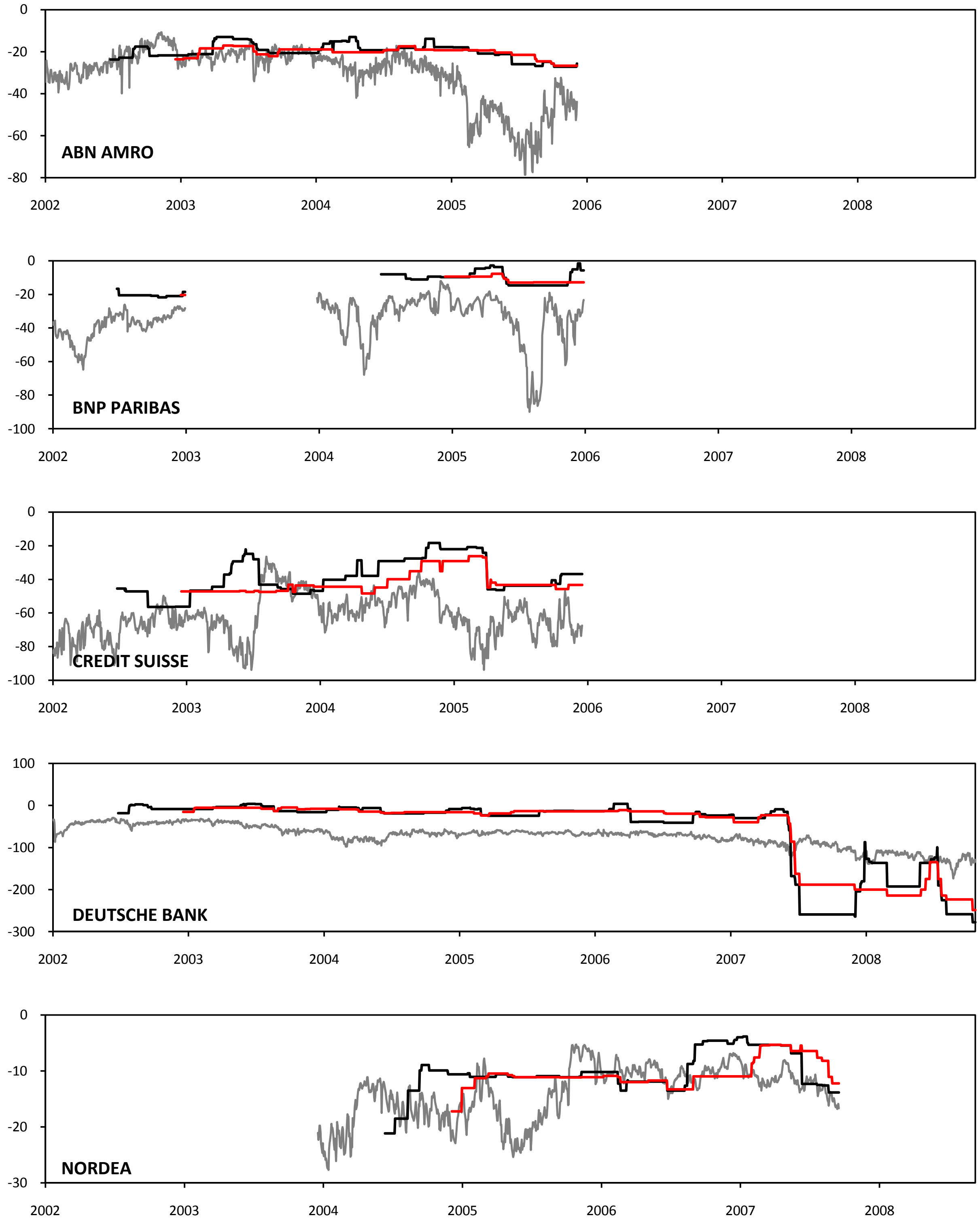
Table 6 | **Parameters ARMA(1,1)-GARCH(1,1) model**

	ABN	BNP	CS	DB	NOR
	Panel A: Parameters in-sample period				
M	**2,21***	**7,05****	**4,19****	**30,86****	-0,12
	(0,81)	**(1,39)**	**(1,44)**	**(1,93)**	(0,29)
Φ	0,43	**0,91****	-0,46	**0,85****	**-0,60****
	(1,97)	**(0,08)**	(0,46)	**(0,07)**	**(0,29)**
Θ	-0,42	**-0,85****	0,56	**-0,80****	**0,70****
	(1,98)	**(0,11)**	(0,42)	**(0,09)**	**(0,26)**
Ω	39,24	59,72	49,49	123,80	0,41
	(78,33)	(49,07)	(41,43)	(362,33)	(0,25)
A	0,03	0,16	0,10	0,01	**0,07****
	(0,06)	(0,11)	(0,05)	(0,03)	**(0,02)**
B	0,69	0,45	**0,80****	0,66	**0,91****
	(0,59)	(0,39)	**(0,13)**	(0,99)	**(0,02)**
Persistence (α+β)	*0,72*	*0,61*	*0,90*	*0,67*	*0,98*
	Panel B: Average parameters out-of-sample period				
M	-35,27	10,61	4,94	45,71	0,31
	(647,46)	(1,11)	(1,57)	(53,84)	(0,33)
Φ	0,08	0,06	0,49	0,78	-0,26
	(0,62)	(0,44)	(0,61)	(0,23)	(0,50)
Θ	-0,13	0,13	-0,41	-0,60	0,36
	(0,62)	(0,45)	(0,62)	(0,29)	(0,53)
Ω	9,24	28,10	94,89	398,90	3,98
	(31,44)	(19,84)	(144,64)	(489,12)	(8,42)
A	0,02	0,15	0,02	0,08	0,07
	(0,07)	(0,08)	(0,07)	(0,13)	(0,07)
B	0,87	0,58	0,60	0,56	0,72
	(0,29)	(0,25)	(0,57)	(0,46)	(0,41)
Persistence (α+β)	*0,89*	*0,73*	*0,62*	*0,64*	*0,79*

Notes: This table presents the parameters of the ARMA(1,1)-GARCH(1,1) model, one of the benchmark models in this thesis. The parameters refer to the coefficients in equations (5) and (6). P&L and VaR figures are estimated in-sample during the first 250 business days (approximately one year) of the sample period. Panel A reports the parameters of the in-sample model. After this period, rolling out-of-sample P&L and VaR figures are estimated using a moving window of 250 preceding business days. Thus, during the out-of-sample period each day new parameters are calculated to produce these figures. Panel B reports the *average* parameters of the out-of-sample models. Standard deviations are in parentheses. The last row of each panel shows the persistence of the models. All statistically significant results are in bold type, using * to denote significance at the 5% level and ** to denote significance at the 1% level.

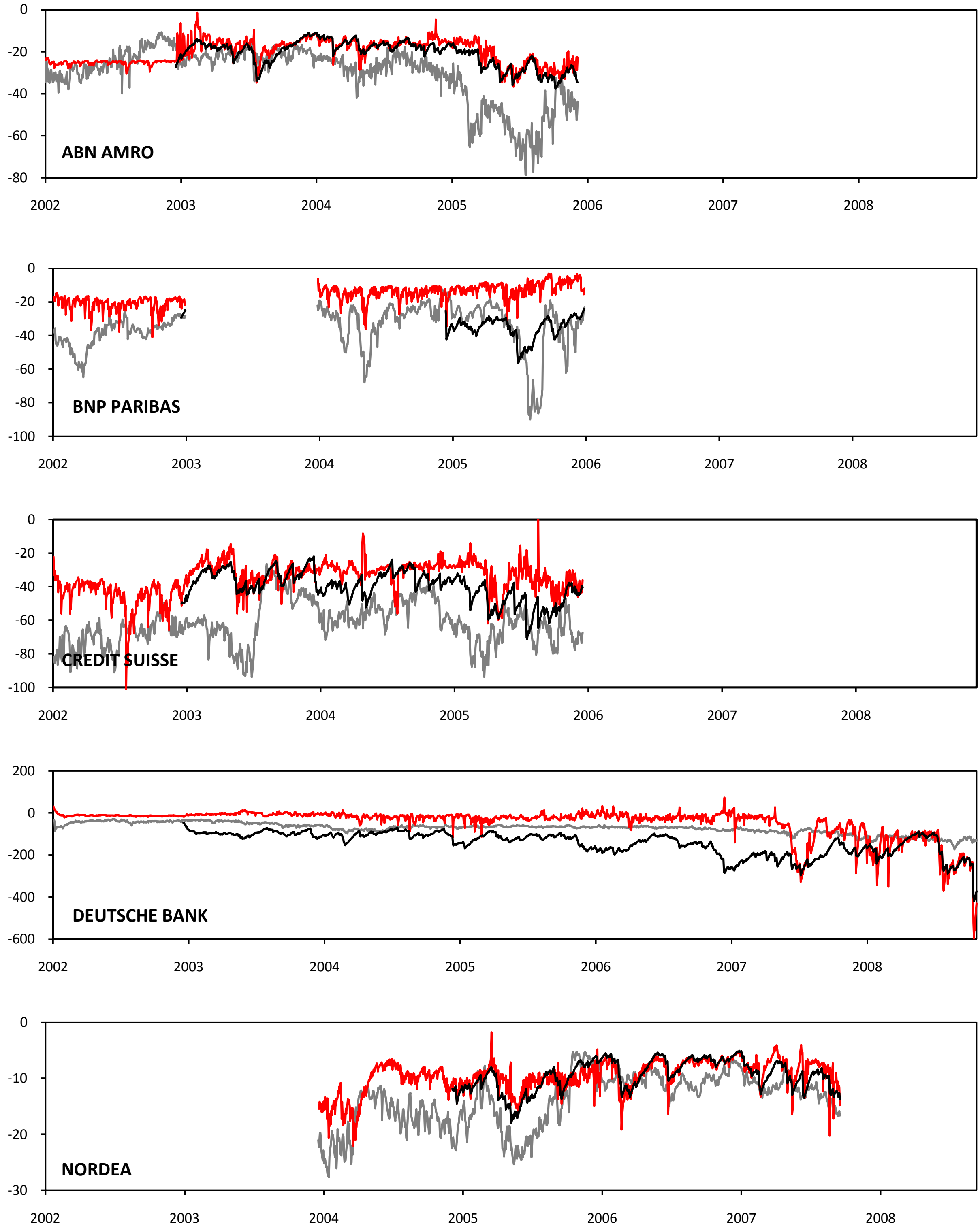
The last benchmark model is the RiskMetrics model with $\lambda = 0{,}94$. The results, depicted in panel E of **TABLE 5**, are mixed. The model yields less exceptions than expected for two banks and more for the other three banks. Also, the models of the latter three banks are accepted based on unconditional coverage tests. For Deutsche Bank the violation rate is so low that the model is just rejected and for BNP Paribas no test statistic is defined. None of the banks can be judged based on conditional coverage tests.

Figure 5a | **Benchmark models: 125-day and 250-day historical simulation models**



Notes: This figure shows for every bank its daily Value-at-Risk (grey line) and two benchmark models. The black line represents a 125-day historical simulation model and the red line a 250-day historical simulation model. Time (in years) is displayed on the horizontal axis. Value-at-Risk is displayed on the vertical axis, in millions of CHF for Credit Suisse and in millions of EUR for the other banks. The total sample period ranges from January 2002 through December 2008. However, not every bank in the sample publishes graphs necessary to extract data throughout the whole sample period. ABN AMRO and Credit Suisse provide graphs in the annual reports of 2002, 2003, 2004 and 2005. The sample period of BNP Paribas is the same, but no graph is published in the annual report of 2003. Nordea presents graphs in 2004, 2005, 2006 and 2007. Deutsche Bank publishes graphs during the entire sample period.

Figure 5b | **Benchmark models: ARMA(1,1)-GARCH(1,1) model and RiskMetrics model**



Notes: This figure shows for every bank its daily Value-at-Risk (grey line) and two benchmark models. The red line represents an ARMA(1,1)-GARCH(1,1) process based on the time series of profits and losses. Using this model, estimates are produced in-sample during the first 250 days. After this time frame, rolling out-of-sample estimates are calculated. The black line represents a RiskMetrics model with $\lambda = 0{,}94$. Time (in years) is displayed on the horizontal axis. Value-at-Risk is displayed on the vertical axis, in millions of CHF for Credit Suisse and in millions of EUR for the other banks.

Plots of the time series (black lines) in **FIGURE 5B** reveal that the RiskMetrics model often produces larger Value-at-Risk estimates (in absolute terms) than the ARMA-GARCH model, hereby yielding less exceptions.

We can summarize that the coverage of the benchmark models is constantly larger than the coverage of the banks' internal models, except for the RiskMetrics model. The 250-day historical simulation model and the RiskMetrics model, both exclusively based on the profit and loss series of the banks, produce better Value-at-Risk estimates than the banks' complicated internal models.

4.3 DEGREE OF RISK DEVIATION

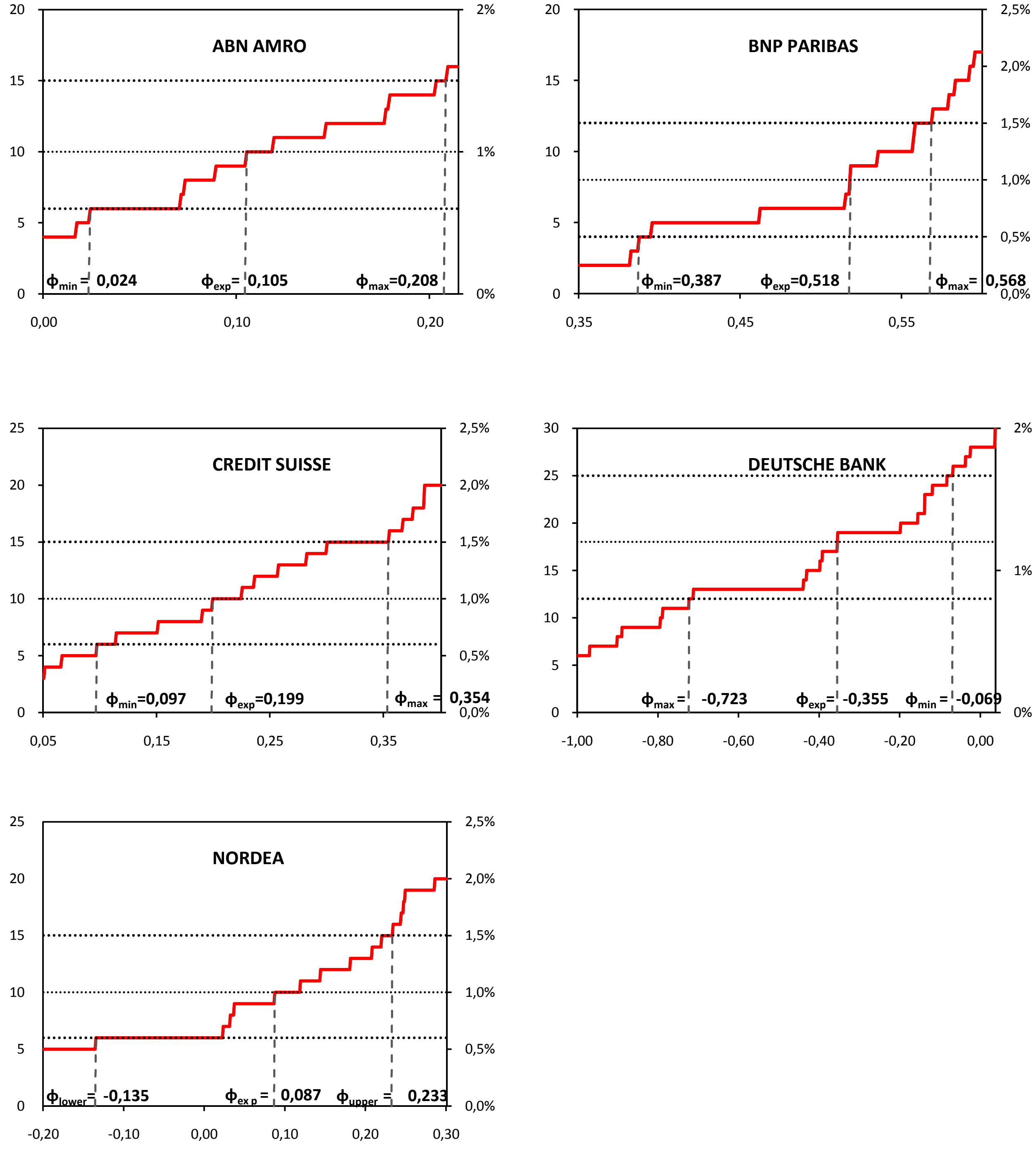
In the previous section we have seen that all of the banks in the sample overstate or understate their risks to a certain degree (not statistically significant each time). Thus, banks consistently provide Value-at-Risk estimates that are too conservative or too tight. However, over- or understatement of risks is not necessarily intended. It might also be the result of mere coincidence. As explained in chapter 3 it is possible to quantify the extent to which banks under- or overstate their risks, using φ_{min}. The resulting graphs are depicted in **FIGURE 6**. Several characteristic points are marked and mentioned in the graphs. The upper and lower dotted lines represent critical values based on a confidence level of 10%. An exception rate of exactly 1% (which implies a perfect model) is marked by the dashed line in the middle. Several remarks concerning the graphs are in order.

First, it should be noted that only Nordea has an acceptable risk model based on the data at hand. The models of this banks reaches the lower bound of the non-rejection zone, but strictly speaking this might be coincidence. Thus, ultimately more exceptions are needed to qualify as a perfect model.

Second, regarding three other banks (ABN AMRO, BNP Paribas and Credit Suisse) risk overstatement is present, but the amount varies across the banks. At a minimum the risk overstatement of the bank with the smallest overstatement, ABN AMRO, amounts to 2,5%.[24] Thus, the disclosed VaR is 2,5% larger than the 'true' VaR. The largest overstatement can be attributed to BNP Paribas and amounts to 63,1% (again at a minimum). The overstatement of Credit Suisse equals 10,7% (again at a minimum).

[24] The percentual (minimum) risk over- or understatement is equal to φ_{min}/(1-φ_{min}), where φ_{min} can be obtained from Figure 5. Thus, the degree of over- or understatement is expressed as a percentage of the 'true' VaR.

Figure 6 | **Risk deviation**



Notes: The horizontal axis in this figure displays a risk deviation coefficient φ, based on the risk overstatement factor of Pérignon *et al.* (2008, p. 791). The left vertical axis displays the number of exceptions and the right vertical axis the corresponding exception rate. The upper and lower dotted lines represent critical values based on a confidence level of 10%. An exception rate of exactly 1% (which implies a perfect model) is marked by the dashed line in the middle. In this paper the upper and lower critical values of φ as well as the 'expected' φ in case of a perfectly specified model are relevant. We define φ_{min} als the upper limit in case of risk understatement (Deutsche Bank) or as the lower limit in case of risk overstatement (ABN AMRO, BNP Paribas and Credit Suisse). The opposite definition is true for φ_{max}. Consequently, in either case $|\varphi_{min}|$ is the smallest value of φ in order to reach the non-rejection zone. Because no risk under- or overstatement is present in the case of Nordea we refer to the boundaries of the non-rejection zone as φ_{lower} and φ_{upper}.

Third, based on our sample Deutsche Bank understates its risk. The number of exceptions (28) lies just above the upper bound of the non-rejection zone (25) in **FIGURE 6**. At a minimum the risk understatement amounts to 6,5%. Stated otherwise, the disclosed VaR is equal to 93,5% of the 'true' VaR. This is a remarkable result, because the supervisor discourages the understatement of risks by means of equation (1) and the penalty scheme depicted in **TABLE 1**.

We can summarize that the direction and degree of risk deviation both vary considerably across the banks. It might be interesting to note that we only have data covering the years 2007 and 2008 ('crisis years') for Deutsche Bank and Nordea. Neither of these banks overstates its risk, unlike the other three banks (for which we have data up to and including 2005). Furthermore, only BNP Paribas shows a large degree of risk overstatement, but this is also the bank with the shortest sample period. The 'data effect' is strengthened if we consider the remaining banks' annual reports of 2008. For example, BNP Paribas and Credit Suisse report 7 and 25 exceptions, respectively, indicating that the financial crisis also affects the models of these banks. Unfortunately, these banks ceased publishing backtesting graphs after 2005. As a consequence, no data could be extracted.

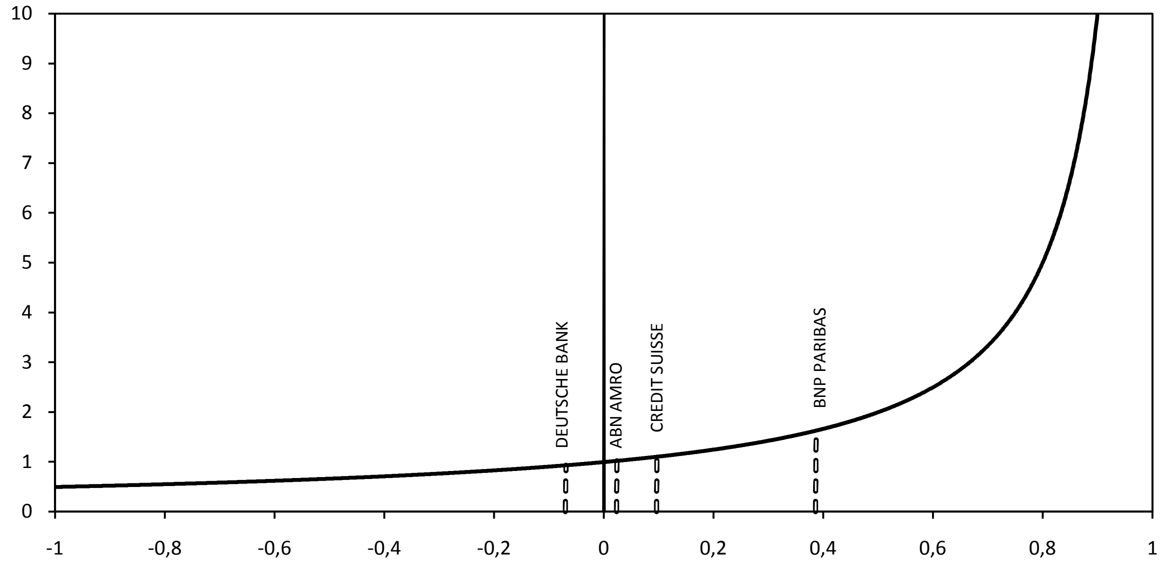
Another interesting aspect relates to the measurement method. According to **TABLE 3** Deutsche Bank is the only bank measuring its risks by means of Monte Carlo simulation (instead of historical simulation). It is also the only bank characterized by significant risk understatement. Although no general conclusions can be made from this single observation, it might indicate that there exists a connection between measurement method and model performance.

4.4 COSTS ASSOCIATED WITH WRONG SPECIFICATION

As explained in chapter 3, costs might be involved with a incorrectly specified Value-at-Risk model. **FIGURE 7** depicts the relationship between the minimum risk deviation coefficient φ_{min} (shown on the horizontal axis) and the ratio MRC_t^{RD}/MRC_t (shown on the vertical axis). We see that the ratio MRC_t^{RD}/MRC_t - and with it the excess market risk charge, see equation (14) - increases exponentially with the risk deviation coefficient: a minimum risk deviation coefficient larger than zero implies a positive excess market risk charge and vice versa.[25]

[25] The result for Nordea is not shown in this graph, because the model of this bank is correctly specified.

Figure 7 | **Monetary effects of wrong model specification**



Notes: The horizontal axis in this figure displays the minimum risk deviation coefficient, φ_{min}, based on the risk overstatement factor of Pérignon *et al.* (2008, p. 791). The vertical axis displays the ratio MRC^{RD}/MRC, where MRC^{RD} is the market risk charge given a too conservative or too tight model and MRC is the market risk charge given a correctly specified model. Thus, the ratio MRC^{RD}/MRC - and with it the excess market risk charge, see equation (14) - increases exponentially with the risk deviation coefficient: a minimum risk deviation coefficient larger than zero implies a positive excess market risk charge and vice versa. The result for Nordea is not shown in this graph, because the model of this bank is correctly specified.

Table 7 | **Example: annual monetary costs of risk overstatement**

	ABN	BNP	CS
Return on assets (ROA)	0,50%	-	0,44%
Disclosed market risk charge ($\mathbf{MRC^{RD}}$)	EUR 480,96	-	CHF 1.119,36
Proper market risk charge (MRC)	EUR 469,42	-	CHF 1010,78
Excess market risk charge (Excess MRC)	EUR 11,54	-	CHF 108,58
MONETARY COSTS (MC)	EUR 0,06	-	CHF 0,48

Notes: This table depicts the monetary costs resulting from risk overstatement. All monetary amounts are in millions. MRC is calculated using equation (13). Subsequently, the excess market risk charge can be obtained using equation (14). If we multiply the excess market risk charge by the opportunity costs of capital – approximated by ROA – that are in force for the different banks, we obtain the monetary costs of risk overstatement (see equation (15)). Results are shown for two of the three risk overstating banks in the sample; necessary data are lacking for BNP.

Ultimately, we can calculate the monetary costs of risk overstatement for the three banks with conservative Value-at-Risk models, using equations (14) and (15).[26] As an example, the results of the calculations for the year 2005 (the last year the risk overstating banks provide the necessary data) are shown in **TABLE 7**. Obviously, it pays to maintain a proper Value-at-Risk model, but the monetary costs of risk overstatement in 2005 are not as large as may be expected in

[26] As mentioned earlier, the calculation of potential ‘benefits’ of risk understatement (in the case of Deutsche Bank) is useless, because these ‘benefits’ are only temporary. Eventually, a bank understating its risks will be penalized. This will lead to a higher market risk charge than necessary, hereby inducing costs.

advance. In fact, in the context of the banks' annual reports these opportunity costs of capital are negligible.

CHAPTER FIVE

CONCLUDING REMARKS

In this study, we have extended the field of empirical research with an analysis of the quality of banks' internal Value-at-Risk models, using a sample of five large European banks. The results indicate that these internal models are often not very accurate; three sample banks overstate their risks and one understates its risks. Also, several benchmark models (e.g. a RiskMetrics model) yield better Value-at-Risk estimates than the internal models.

If the results of this study are compared to prior research we see some striking differences. Indeed, both Berkowitz and O'Brien (2002) and Pérignon *et al.* (2008) find that banks' internal models provide inaccurate risk estimates, but they also show that this is due to conservativeness of the banks in their sample. On the other hand, we have shown that our findings are mixed. Indeed, we have confirmed that models are incorrectly specified, but not only because banks overstate their risks. Only one bank (also the one with the shortest time series) is characterized by severe risk overstatement. In one case, even risk understatement is present.

Considering our and prior empirical findings the question arises whether (internal) Value-at-Risk models are appropriate instruments to measure market risk. Apparently, banks are not able or willing to construct Value-at-Risk models such that they yield accurate estimates of market risk. The backtesting procedure proposed by the Basel Committee on Banking Supervision does not provide the incentive to solve this problem either. In line with Lucas (2001) we argue that under the current penalty scheme banks might be tempted to use a model that is not correctly specified.[27] A stricter penalty scheme might

[27] As discussed in chapter 2, according to Lucas (2001) this might eventually lead to deliberate risk understatement.

remove some of these moral hazard problems. If model failure is still present after the implementation of such a scheme, then it is obvious that capital requirements should not be based on the internal models in their current appearance. We have shown that simple Value-at-Risk models based on the profit and loss series of the banks outperform the complicated internal models. In their conclusion, Berkowitz and O'Brien (2002, p. 1110) state that "time-series models may have advantages in forecasting and as a tool for identifying the shortcomings of the structural model [the internal model, SB]". Based on our results, we underline this remark.

The incorporation of the global financial crisis in the data set has contributed to the results as they are, although we only have data covering the crisis years for Deutsche Bank and Nordea. Large losses have led to an unprecedented number of exceptions for these banks. The other sample banks show similar numbers. Obviously, the banks' internal models were unable to capture the dynamics of the crisis. This is confirmed by the results of Pérignon and Smith (forthcoming). Thus, an additional recommendation might be to expose the internal Value-at-Risk models to extreme scenarios (stress testing). The usefulness of this approach has also been recognized by the Basel Committee on Banking Supervision (2009). As a response to the financial crisis the Committee plans to modify the regulatory framework by incorporating several stress testing requirements. These modifications should take effect by the end of 2010. In the Netherlands, the Advisory Committee on the Future of Banks in the Netherlands (2009) released the document Restoring Trust. This document deals with a number of recommendations in order to restore confidence in the Dutch banks. Among other things, the Advisory Committee (2009, p. 31) states that "the models used by banks to determine market risks need to take more account of events that are extremely unlikely but that would have a major impact if they do happen". The recommendations are implemented as of January 1, 2010.

We can conclude that some modifications of the existing framework might lead to a better control of market risks. First, calculating Value-at-Risk based on a reduced-form model next to the structural model might provide better insight into the real market risk. Second, the problems accompanying the determination of proper market risks have become the subject of attention of the regulator during the global financial crisis. Revisions to fight these problems, most notably the implementation of a comprehensive stress testing program, are in progress. In the years to come it will become clear whether the latter measures turn out to be effective.

REFERENCES

Advisory Committee on the Future of Banks in the Netherlands. (2009). *Restoring trust*. Nederlandse Vereniging van Banken, Amsterdam.

Basel Committee on Banking Supervision. (1988). *International convergence of capital measurement and capital standards.* Bank for International Settlements, Basel.

Basel Committee on Banking Supervision. (1996). *Amendment to the capital accord to incorporate market risks*. Bank for International Settlements, Basel.

Basel Committee on Banking Supervision. (2004). *International convergence of capital measurement and capital standards: a revised framework*. Bank for International Settlements, Basel.

Basel Committee on Banking Supervision. (2006). *International convergence of capital measurement and capital standards: a revised framework: comprehensive version*. Bank for International Settlements, Basel.

Basel Committee on Banking Supervision. (2009). *Revisions to the Basel II market risk framework.* Bank for International Settlements, Basel.

Berkowitz, J., & O'Brien, J. (2002). How accurate are Value-at-Risk models at commercial banks?, *Journal of Finance*, *57*(3), 1093-1111.

Berkowitz, J., Christoffersen, P.F., & Pelletier, D. (2008). *Evaluating Value-at-Risk models with desk-level data.* Working paper. McGill University, Montreal.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, *31*, 307-327.

Box, G. P. E., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control, second edition*. San Franscisco: Holden-Day.

Campbell, S.D. (2005). *A review of backtesting and backtesting procedures*. Working paper. Federal Reserve Board, Washington, D.C.

Christoffersen, P.F. (2003). *Elements of Financial risk management*. Amsterdam: Academic Press.

Deutsche Bank. (2007). *Annual report*. Retrieved from http://www.db.com/ir/en/content/reports_2007.htm

Hendricks, D. (1996). Evaluation of Value-at-Risk models using historical data, *Federal Reserve Bank Economic Policy Review*, *April*, 39-69.

Hendricks, D., & Hirtle, B. (1997). Bank capital requirements for market risk: the internal models approach, *Federal Reserve Bank Economic Policy Review*, *December*, 1-12.

J.P. Morgan. (1996). *RiskMetrics*. Technical document. New York.

Jorion, P., & Taleb, N. (1997). The Jorion-Taleb debate. *Derivatives Strategy*, *2(4)*. Retrieved from http://www.derivativesstrategy.com/magazine/archive/1997/0497fea2.asp

Jorion, P. (2007). *Value at Risk: the new benchmark for managing financial risk, third edition*. New York: McGraw-Hill.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models, *Journal of Derivatives*, 3, 73-84.

Lopez, J.A. (1999). Regulatory Evaluation of Value-at-Risk Models, *Journal of Risk*, *1*, 37-64.

Lucas, A. (2001). Evaluating the Basle guidelines for backtesting banks' internal risk management models, *Journal of Money, Credit, and Banking*, *33*(3), 826-846.

Pérignon, C., Deng, Z.Y., & Wang, Z.J. (2008). Do banks overstate their Value-at-Risk?, *Journal of Banking & Finance*, *32*, 783-794.

Pérignon, C., & Smith, D.R. (Forthcoming). The level and quality of Value-at-Risk disclosure by commercial banks, *Journal of Banking & Finance*.

Furthermore, it should be noted that we have used the annual reports of ABN AMRO, BNP Paribas, Credits Suisse, Deutsche Bank and Nordea (years 2001 up to and including 2008) to compile our data set and to look up additional information.

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my thesis supervisor, Remco Zwinkels, for the advice and guidance I received in completing my master thesis. Furthermore, I would like to thank the co-reader for taking the time to read my work. And finally, I am particularly grateful to my parents and their partners for their support and confidence in the course of completing my studies.

Sander van den Boogaart