

Optimal simultaneous validation tests of default probabilities, dependencies, and credit risk models

Uwe Wehrspohn

This version: July 15, 2004

Department of Economics, Heidelberg University, Grabengasse 14, 69117 Heidelberg, Germany (e-mail: uwe.wehrspohn@urz.uni-heidelberg.de).

Correspondence address: Center for Risk & Evaluation, Berwanger Strasse 4, 75031 Eppingen, Germany (e-mail: wehrspohn@risk-and-evaluation.com).

Abstract. We provide a general, model-independent approach to the construction of optimal simultaneous validation tests of credit default probabilities, dependencies between creditors, and credit risk models that maximize the power of test for any given portfolio-size and number of periods of data available. Results can be used to validate banks' estimates of rating default probabilities, correlations and choice of credit risk models in the Basel II supervisory review process. Example-analyses are given for the generalized asset value model.

Key words: rating validation, backtesting, model validation, banking regulation

JEL Classification: C5, C52

1 Introduction

Clients' default probabilities, dependencies between clients and the choice of the particular unifying credit risk model are the main drivers of the credit risk found in bank portfolios and in the banking sector in general. Errors in the estimation of these key inputs may lead to considerable misconceptions of the overall credit risk a financial institution is exposed to and, in turn, also lead to insufficient action by risk managers as well as regulators.

Therefore, risk managers and regulators are vitally interested in having at their disposal a technology to validate key risk drivers as a cornerstone of a consolidated credit risk management process.

Moreover, in most approaches to model portfolio credit risk the estimations of clients' default probabilities and of the size of dependencies between clients are interrelated in the sense that dependencies result solely from an incomplete knowledge of the factors that commonly influence clients' default probabilities. Assuming increasing ex ante information about these systematic risk factors entails a refined estimation of default probabilities and usually decreases dependencies between defaults of clients. This implies that default probabilities, dependencies, and their interrelation as described by the credit risk model applied should not be assessed in isolation, but in their modelled connection.

In this article, we provide a general, model-independent approach to the construction of optimal simultaneous validation tests of credit default probabilities, dependencies between creditors, and credit risk models. The resulting tests maximize the power of test for any given number of clients and number of periods of data available. They, thus, also represent a benchmark for the maximal quality of validation tests a bank can reach if the amount of data at disposition is limited.

The paper is organized as follows: section 2 gives a brief overview about the literature, section 3 provides the general validation approach, section 4 applies the approach to the generalized asset value model and serially independent defaults and gives quantitative examples. Section 5 concludes.

2 Brief Review of the Literature

Jose Lopez and Marc Saldenberg (2000) were among the first who suggested to consider credit portfolio model structures to facilitate validation of model inputs, outputs, and of the model itself. In order to compensate for the sparsity of historical data, they proposed a bootstrap approach. Making the implicit assumption that defaults are independent, large numbers of smaller portfolios were sampled from a large portfolio, and model results were compared to actual portfolio performance so that numerous standard tests could be applied. If defaults happen to be dependent, though, as usually has to be accounted for in real world situations, the Lopez- and Saldenberg-approach becomes invalid.

Taking up an idea of Jeremy Berkowitz (2001), Hergen Frerichs and Günter Löffler (2003) suggest to apply the estimated cumulative distribution function \hat{F} of portfolio defaults to the number of observed defaults x in order to transform them to the unit interval and then in a second step to further transform them to the real axis by the inverse standard normal cumulative distribution function Φ^{-1} . They then test the hypothesis that $y := \Phi^{-1}(\hat{F}(x))$ be standard normally distributed against the alternative that y follows a specific non-standard normal distribution that depends on the quantity to be tested.

Albeit their results show large imprecisions under the null which cannot be due to simulation uncertainty, the authors do not consider the impact of some misspecifications of the constructed test. In particular, the distribution of portfolio defaults is in general not absolutely continuous with respect to the Lebesgue measure. y can, therefore, not be standard normally distributed even under the null. Hence, the test-statistic cannot be χ_n^2 -distributed under the null, as the

authors state without proof. Own simulation exercises have shown that the true distribution of the test-statistic under the null is far longer tailed than the χ_n^2 -distribution and that the misspecification of the correct distribution leads to a relative overstatement of the power of test of up to 53%.

Moreover, even if the distribution of portfolio defaults were continuous with respect to the Lebesgue measure, y would not be normally distributed under the alternative so that the test is suboptimal even in the limiting case.

3 General Validation Approach

In order to be able to describe our general validation approach, we define some notation and make the following assumptions:

- Default data is available for periods $t = 1, \dots, T$.
- In period t clients are indexed $i = 1, \dots, n_t$ for $t = 1, \dots, T$.
- Let H and A denote the hypothesis and alternative, respectively.
- Under the hypothesis, in period t clients have default probabilities $p_{1t}^H, \dots, p_{n_t t}^H$ for $t = 1, \dots, T$.
- Under the alternative, in period t clients have default probabilities $p_{1t}^A, \dots, p_{n_t t}^A$ for $t = 1, \dots, T$.
- Under the hypothesis, default distributions are determined by credit portfolio model M^H .
- Under the alternative, default distributions are determined by credit portfolio model M^A .
- Under the hypothesis, portfolio model M^H can be fully parameterized through clients' default probabilities and a list D^H of known parameters that comprises all dependencies between clients.
- Under the alternative, portfolio model M^A can be fully parameterized through clients' default probabilities and a list D^A of known parameters that comprises all dependencies between clients.
- Let $\mathbf{1}_{it} = 1$ if client i defaults in period t , and $\mathbf{1}_{it} = 0$ otherwise for $i = 1, \dots, n_t$ and $t = 1, \dots, T$.
- Let $x_{it} \in \{0, 1\}$ for $i = 1, \dots, n_t$ and $t = 1, \dots, T$.
- Let

$$p_{[x_{11}, x_{21}, \dots, x_{n_T T}]}^I \tag{1}$$

$$:= \mathbb{P}_I\{\mathbf{1}_{11} = x_{11}, \dots, \mathbf{1}_{n_T T} = x_{n_T T}\}, \tag{2}$$

$$:= \mathbb{P}\{\mathbf{1}_{11} = x_{11}, \dots, \mathbf{1}_{n_T T} = x_{n_T T} \mid M^I, D^I, p_{11}^I, \dots, p_{n_T T}^I\}, \tag{3}$$

$I = A, H$, be the probability to observe the event $\mathbf{1}_{11} = x_{11}, \dots, \mathbf{1}_{n_T T} = x_{n_T T}$ under the hypothesis or alternative, respectively.

- Define the test-statistic T as

$$T(x_{11}, \dots, x_{n_T T}) := \frac{p_{[x_{11}, \dots, x_{n_T T}]}^A}{p_{[x_{11}, \dots, x_{n_T T}]}^H}. \quad (4)$$

- Let

$$\phi(x_{11}, \dots, x_{n_T T}) := \mathbb{P}\{\text{H is rejected} \mid x_{11}, \dots, x_{n_T T}\}$$

be a randomized test of the hypothesis H against the alternative A .

We can now state our main result.

Theorem 1

Let $\gamma \in [0, 1]$ and $k \in \mathbb{R}^{\geq 0}$ and let $\alpha \in (0, 1)$ be a significance level. Let

$$\phi(x_{11}, \dots, x_{n_T T}) := \begin{cases} 1 & \text{if } T(x_{11}, \dots, x_{n_T T}) > k \\ 0 & \text{if } T(x_{11}, \dots, x_{n_T T}) < k \\ \gamma & \text{if } T(x_{11}, \dots, x_{n_T T}) = k \end{cases} \quad (5)$$

Then γ and k can be chosen so that

$$\mathbb{E}_H \phi(x_{11}, \dots, x_{n_T T}) = \alpha \quad (6)$$

and the test ϕ maximizes the power among all tests of hypothesis H against alternative A .

Proof:

The result follows from Lehmann (1997) chapter 3.2. theorem 1.

□

A number of important implications directly follow from the theorem:

- Theorem 1 is valid for any portfolio models.
- Theorem 1 allows to test all model parameters, for instance default probabilities, individually or simultaneously, even including the model itself. In all cases, dependencies between clients are naturally integrated in the analysis.
- Although usually difficult to calculate analytically, the distribution of the test-statistic T under the hypothesis and the alternative, respectively, can always be approximated by computer simulation. The same is true for the probabilities $p_{[x_{11}, \dots, x_{n_T T}]}^H$ and $p_{[x_{11}, \dots, x_{n_T T}]}^A$.
- The test in theorem 1 maximizes the power, i.e. the probability to correctly reject the hypothesis if the alternative is true, for any given amount of data. This means that it allows for optimal testing even if data is sparse, i.e. if there are only few periods of data and / or few clients.

- Theorem 1 establishes a benchmark that makes evident what power of test is achievable given a certain amount of data.
- It also gives a clue whether the prospective costs of data collection lead to the desired increase in power of test.

4 An Example

To illustrate the approach, we apply theorem 1 to a specific family of credit portfolio models, the generalized asset value model.

The classical asset value model was developed by Stephen Kealhofer, Andrew McQuown, and Oldrich Vasicek (see e.g. Kealhofer (1993)) and by Greg Gupton, Christopher Finger, and Mickey Bhatia (1997) based on a seminal article by Robert Merton (1974) and is amply documented in literature.

The generalized asset value model (for details see Wehrspohn (2002), pp. 112ff., Wehrspohn (2003)) extends the classical model in the way that the normal distribution as asset return distribution is replaced by a general variance mixture of normals. This is an important topic for testing because it can be shown that tail-risk increases for any deviations from the normal distribution as asset return distribution in the generalized asset value model (Wehrspohn (2003)).

To keep the number of parameters as small as possible, we assume that data is available about n clients in periods $t = 1, \dots, T$ independent of t . All clients have default probabilities p in all periods. Defaults are serially independent. Within the same period, asset return correlations ρ between two asset return distributions may be positive and are the same for any two clients and all periods.

In the generalized asset value model, firm i , $i = 1, \dots, n$, defaults in a given period if

$$\sqrt{w} \left(\sqrt{\rho} Y + \sqrt{1 - \rho} Z_i \right) \leq F^{-1}(p) \quad (7)$$

where Y and Z_i , $i = 1, \dots, n$, are independent standard normally distributed random variables, w is a random variable that only takes values on \mathbb{R}^+ and is independent of Y and Z_i , $i = 1, \dots, n$, and F is the cumulative distribution function of $\mathcal{L}(\sqrt{w} \cdot Y)$. Note that Y and w are systematic risk factors common to all clients, while Z_i is individual to client $i = 1, \dots, n$.

The classical asset value model appears as a special case of the generalized asset value model for $w \equiv c$ and some positive constant c .

Let $x_{1t}, \dots, x_{nt} \in \{0, 1\}$ be defined as above and let $\sum_{i=1}^n x_{it} = m_t$. Then it can be shown (Wehrspohn (2002) theorem 11) that for all $t = 1, \dots, T$

$$P_{[x_{1t}, \dots, x_{nt}]} = \int_{\mathbb{R}^+} \int_{\mathbb{R}} \left[\Phi \left(\frac{F^{-1}(p) - \sqrt{w} \cdot \sqrt{\rho} \cdot Y}{\sqrt{1 - \rho}} \right) \right]^{m_t} \cdot \left[1 - \Phi \left(\frac{F^{-1}(p) - \sqrt{w} \cdot \sqrt{\rho} \cdot Y}{\sqrt{1 - \rho}} \right) \right]^{n - m_t} \cdot \phi(Y) dY \cdot W(dw) \quad (8)$$

where Φ is the cumulative standard normal distribution function, ϕ is the standard normal density function, and $W = \mathcal{L}(w)$.

Due to the assumed serial independence of defaults, it follows that

$$p_{[x_{11}, \dots, x_{nT}]} = \prod_{t=1}^T p_{[x_{1t}, \dots, x_{nt}]} \quad (9)$$

with a specific combination of m_t defaults in period $t = 1, \dots, T$.

Let now the model parameters p and ρ and the model itself, i.e. the mixing distribution $W = \mathcal{L}(w)$, depend on the hypothesis and the alternative, respectively, to get p^I , ρ^I and W^I , $I = A, H$. Then the test-statistic T is given as

$$T(x_{11}, \dots, x_{nT}) := \frac{p_{[x_{11}, \dots, x_{nT}]}^A}{p_{[x_{11}, \dots, x_{nT}]}^H}. \quad (10)$$

To be able to produce numerical results, we further specialize the example and choose the symmetric version of the normal inverse Gauss distribution as asset return distribution to get a subfamily of the generalized asset value model. The symmetric normal inverse Gauss distribution has the density

$$\text{nig}(x; \alpha, \delta, \mu) = \frac{\alpha\delta}{\pi} \exp(\delta\alpha) \frac{K_1\left(\alpha\sqrt{\delta^2 + (x - \mu)^2}\right)}{\sqrt{\delta^2 + (x - \mu)^2}} \quad (11)$$

where $x, \mu \in \mathbb{R}$, $\delta \geq 0$, and $K_\lambda(\cdot)$ is the modified Bessel function of the third kind.

The symmetric normal inverse Gauss distribution can be written as a variance mixture of normals with the inverse Gauss distribution as mixing distribution. The inverse Gauss distribution has the density

$$\text{ig}(x; \psi, \chi) = \sqrt{\frac{\chi}{2\pi x^2}} \cdot \exp\left(-\frac{1}{2}\left(\frac{\chi}{x} + \psi x\right) + \sqrt{\psi\chi}\right) \quad (12)$$

with $\chi = \delta^2$ and $\psi = \alpha^2$. Without loss of generality, we can choose the symmetric normal inverse Gauss distribution to have unit variance, i.e. $\alpha = \delta$. δ now is a free shape-parameter that leaves the variance unchanged, but deforms the tails of the distribution. Note that for $\delta \rightarrow \infty$ the normal inverse Gauss distribution converges against the normal distribution. For $\delta < \infty$ the distribution has long tails (Wehrspohn (2002) theorem 4) that turn longer with falling δ . For an extensive discussion of the normal inverse Gauss, the inverse Gauss and other related distributions refer to Eberlein and Prause (1998) and Prause (1999).

For the numerical analyses, we choose as the base case that data is available for 5 periods for a portfolio containing 500 clients. To show the impact of the size of the portfolio and of the length of history of data, we deviate from the base case to histories of 1, 3, 5, 10, and 15 periods of data on the one hand, and to portfolios containing 50, 100, 250, 500, 1000, and an infinite number of clients on the other hand. The significance level (error of the first kind) is kept fixed at 0.05.

For all analyses, the hypothesis is that clients have default probabilities $p = 0.01$, asset return correlations $\rho = 0.1$, and that the asset return distribution is normal. The alternative always only deviates in one parameter-value from the

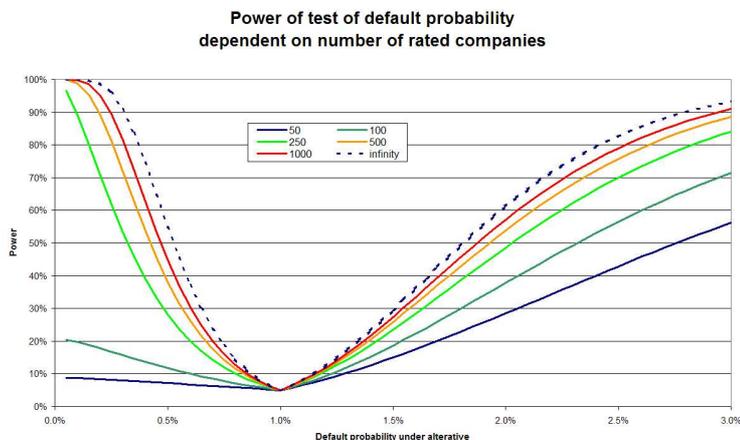


Figure 1: Tests on default probability for varying portfolio sizes

hypothesis. We successively test default probabilities to range from 0.0005 to 0.03, asset return correlations to range from 0 to 0.3, and the shape parameter δ of the normal inverse Gauss distribution to range from 0.5 to 20.

Figure 1 shows the power of tests on deviations from a default probability of 0.01 for portfolio sizes varying from 50 to 1000 and infinitely many clients, respectively. It turns out that in very small portfolios containing less than 250 clients it is very unlikely to detect even larger deviations from the hypothesis. On the other hand, if the portfolio size exceeds 250, the power increases rapidly so that the comparative advantage of large banks over middle-sized banks with 500-1000 clients in one rating class is relatively small.

Figure 2 displays the power of the same test for a portfolio with 500 clients if varying lengths of histories of data are available. It is evident that long histories of data significantly improve the results and sometimes more than doubles the power of test, particularly if the hypothesis is tested against deviations to higher default probabilities. Thus, also for regulatory purposes, it is worthwhile to collect data over longer periods if this is possible.

Figures 3 and 4 show the same analyses for tests on deviations of the asset return correlations from the hypothesis of a correlation of 0.1. Note that it turns out that here deviations from the hypothesis are much more difficult to detect than deviations in default probability. This seems to indicate that in middle-sized portfolios default probabilities are the primary driver of default behavior of clients. This is in line with the observation that large banks have a greater advantage over small and middle-sized banks when asset return correlations are tested as opposed to a test of default probabilities.

Figures 5 and 6 present the results for deviations from the normal distribution as asset return distribution. The choice of the asset return distribution is critical in the generalized asset value model because it can be shown that tail-risk and, thus,

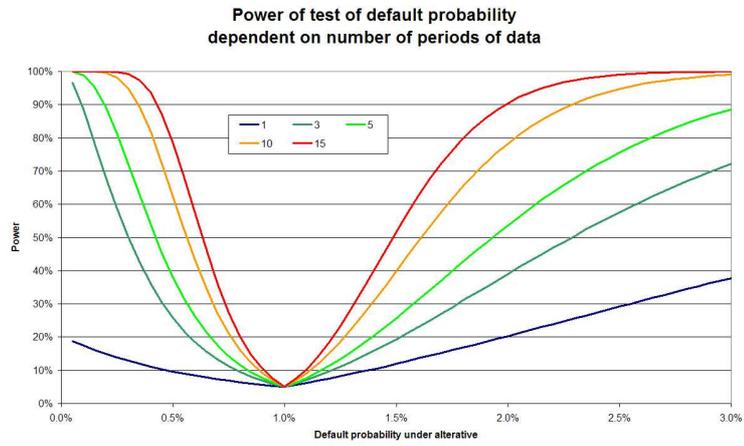


Figure 2: Tests on default probability for varying length of data histories

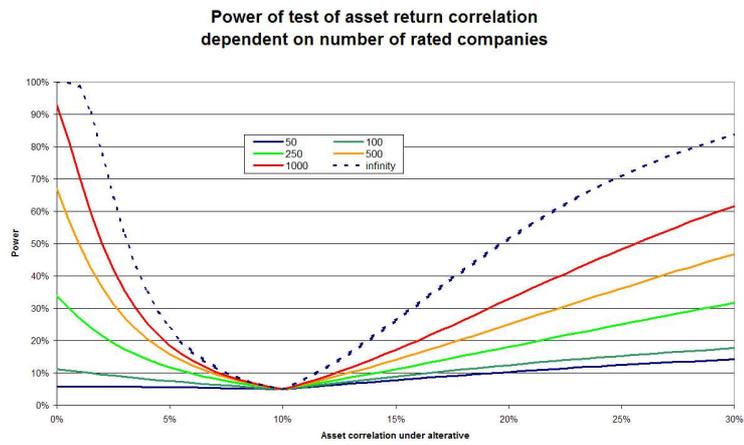


Figure 3: Tests on asset return correlations for varying portfolio sizes

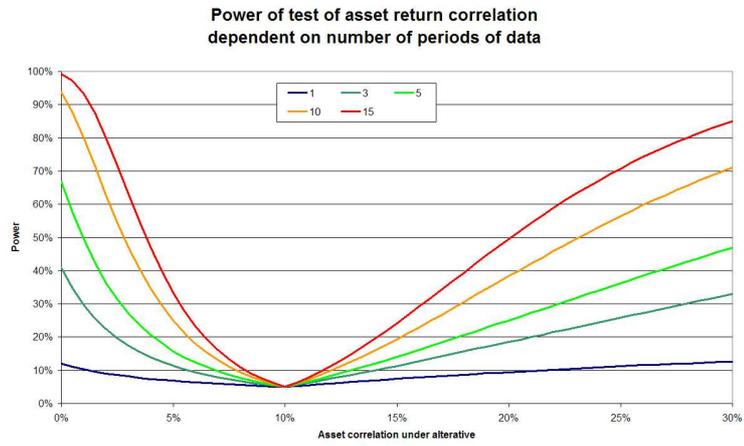


Figure 4: Tests on asset return correlations for varying length of data histories

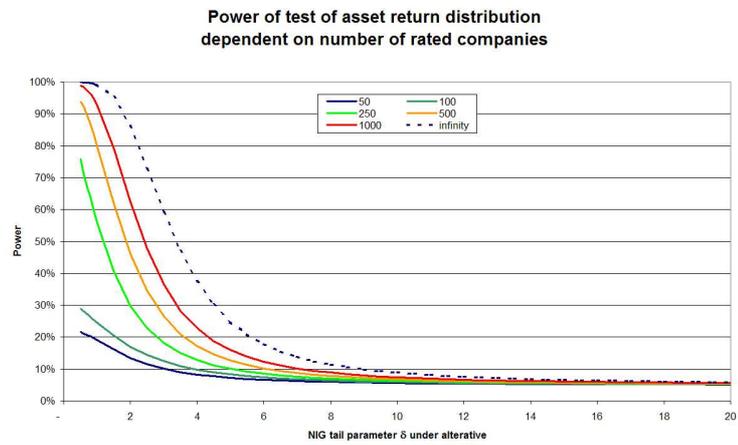


Figure 5: Tests on asset return distributions for varying portfolio sizes

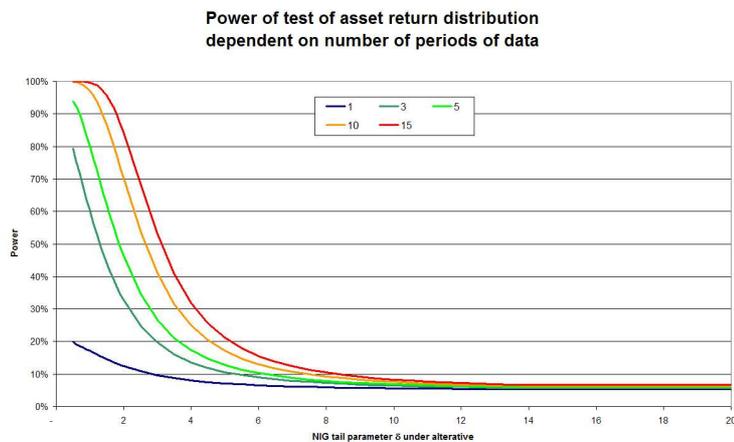


Figure 6: Tests on asset return distributions for varying length of data histories

values at risk and shortfalls at high confidence levels increase for any deviations from the normal distribution as asset return distribution (Wehrspohn (2003)). Note that small deviations from the normal distribution, i.e. values of $\delta \geq 8$ are very difficult to detect even if the portfolio is large and many periods of data are available. Only alternatives that are further away from the hypothesis are more readily detected. Similar to the tests of asset return correlations, here again to have a large portfolio and a long history of data is clearly advantageous.

Finally note that in all cases a power of test above 40% can only be reached if the alternative deviates from the hypothesis to a relatively large extent, albeit the fact that the test under discussion maximizes the power and is already a best test. This indicates that it is under all circumstances paramount for banks and regulators to pool data, if possible, to get large portfolios and to collect long histories of default observations.

5 Conclusion

We have defined a general, model-independent framework that allows banks and regulators to construct optimal validation tests for their credit risk models and risk parameters. We have applied the approach to the generalized asset value models and have given numerical examples for the performance of tests on default probabilities, asset return correlations, and the asset return distribution.

The results have shown that it is relatively easy to test on deviations of default probabilities from a given hypothesis as compared to asset return correlations and asset return distributions. However, high detection probabilities, if the alternative is true, can only be obtained if the alternative is far away from the hypothesis or if there are many periods of data available for a large portfolio. This shows that data consolidation is one of the most important tasks for banks and regulators.

References

- [1] Berkowitz, Jeremy, 2001, "Testing density forecasts with applications to risk management," *Journal of Business & Economic Statistics* 19, pp. 465-474
- [2] Eberlein, Ernst and Karsten Prause, 1998, "The generalized hyperbolic model: financial derivatives and risk measures," Working paper Nr. 56, Freiburg University
- [3] Frerichs, Hergen and Günter Löffler, 2003, "Evaluating credit risk models using loss density forecasts," Working Paper
- [4] Gupton, Greg M., Christopher C. Finger, Mickey Bhatia, 1997, "Credit Metrics - Technical Document," Working Paper, JP Morgan
- [5] Kealhofer, Stephen, 1993, "Portfolio management of default risk," Working paper, KMV Corporation, <http://www.kmv.com>
- [6] Lehmann, Erich Leo, 1997, "Testing Statistical Hypotheses," 2nd edition, Springer-Verlag, New-York
- [7] Lopez, Jose and Marc Saldenberg, 2000, "Evaluating credit risk models," *Journal of Banking and Finance* 24, pp. 151-165
- [8] Merton, Robert C., 1974, "On the pricing of corporate debt: the risk structure of interest rates," *Journal of Finance* 39, pp. 449-470
- [9] Prause, Karsten, 1999, "The generalized hyperbolic model: estimation, financial derivatives, and risk measures," Dissertation Freiburg University
- [10] Wehrspohn, Uwe, 2002, "Credit Risk Evaluation," Dissertation Heidelberg University, <http://www.wehrspohn.de>
- [11] Wehrspohn, Uwe, 2003, "Generalized Asset Value Credit Risk Models and Risk Minimality of the Classical Approach," Working Paper, <http://www.wehrspohn.de>