Measuring model risk

Philipp Sibbertsen
Department of Economics, Leibniz University Hannover, Königsworther Platz 1, 30167 Hannover, Germany; email: sibbertsen@statistik.uni-hannover.de

Gerhard Stahl
Talanx AG, Riethorst 2, 30659 Hannover, Germany; email: gerhard.stahl@talanx.com
and
Ulm University, Germany

Corinna Luedtke
Department of Economics, Leibniz University Hannover, Königsworther Platz 1, 30167 Hannover, Germany; email: luedtke@statistik.uni-hannover.de

Model risk as part of operational risk is a serious problem for financial institutions. As the pricing of derivatives as well as the computation of the market or credit risk of an institution depend on statistical models, the application of a wrong model can lead to a serious over- or underestimation of the institution's risk. Because the underlying data-generating process is unknown in practice, evaluating the model risk is a challenge. So far, definitions of model risk have been either application-oriented, including risk induced by the statistician rather than by the statistical model, or so research-oriented as to prove too abstract to be used in practice. They are especially prone not to be data driven. We introduce a data-driven notion of model risk that includes the features of the research-oriented approach, extending it by a statistical model-building procedure and thus compromising between the two definitions at hand. We further suggest the application of robust estimates to reduce the model risk, and advocate the application of stress tests with respect to the valuation of the portfolio.

1 INTRODUCTION

Model risk emerged as a new risk category with the advent of elaborate and complex mathematical and statistical models during the late 1970s, when Black, Scholes and Merton laid the path for a new field of science – mathematical finance. It is under the focus of various stakeholders of so-called internal models: regulators, rating agencies, investors, shareholders, bondholders, chartered accountants and the firm's board. Therefore, model risk can hardly be overestimated. A quote from a sector report of the UniCreditGroup (2008) states that some investors "tend to apply an across-the-board discount of about 20% to the published numbers"; this tendency is said to be due to investor skepticism and lack of transparency, giving the reader a sense of its economic relevance.

The urge for better models stemmed from a riskier environment for financial institutions after the breakdown of the Bretton Woods system and the steadily

FIGURE 1 Risk measurement procedure (Cont *et al* (2007)).

$$
\begin{tikzcd}
& D_\rho \arrow[dr, "\rho"] & \\
X \arrow[ur, "M"] \arrow[rr, "\hat{\rho} = \rho \circ M"'] & & \mathbb{R}
\end{tikzcd}
$$

growing interdependence of the world economy. Financial institutions were gradually exposed to higher risks and faced new risk categories. At the same time, the huge increase in computer capacity and capabilities led to better possibilities for calculating these models. An overview of the technical progress and its implications for financial markets is given in Marshall (2000). Alexander (2003) and Padoa-Schioppa (2004) show the rationale of regulation, especially of so-called internal models for banks. In 1998, regulatory authorities acknowledged these models for regulatory purposes. Since then they have been widespread in practice.

However, a model is only capable of capturing features of reality and is subject to the modeler's preferences and abilities. Even highly sophisticated models capturing stylized facts such as volatility clustering or long memory are confronted with mapping errors when implemented in practice. The role of mapping errors is depicted in Figure 1, where X denotes the data set, M is the estimator of the loss distribution, D_ρ is the domain and ρ is the risk measure. The combination $\hat{\rho} = \rho \circ M : X \to \mathbb{R}$ allows for the direct computation of a risk measure estimate.

The final output of an internal model is a realization x_{t+h} associated with a forecast distribution:

$$L(\pi_{t+h} \mid I_t)$$

and:

$$L(\pi_{t+h} \mid I_t, Z_t)$$

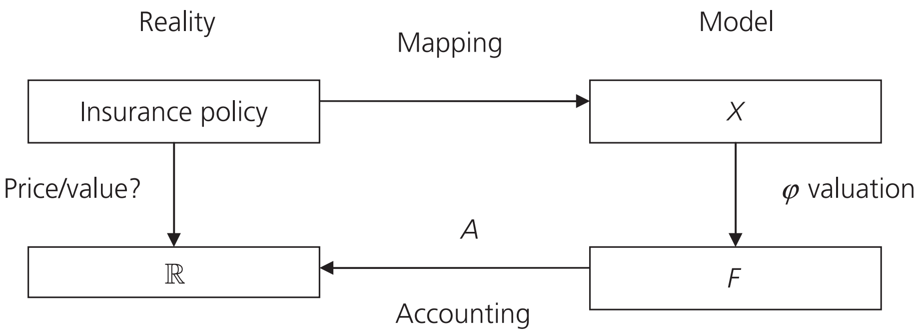
respectively, where π_{t+h} denotes the variable of interest, typically a portfolio that is held over a time span h. The forecast is drawn on two different types of information sets. The first, I_t, comprises a history of risk factors, eg, a time series of bond prices. The second, Z_t, which is typically used by insurers under Solvency II, is of managerial type, eg, management rules for splitting losses between shareholders and policyholders.

But this is only half of the output. The other is related to the valuation of π_{t+h}:

$$v(\pi_{t+h})$$

Figure 2, taken from Baumgartner *et al* (2004), shows the interplay between the model and the various valuations. From this representation one concludes that in the insurance industry mark-to-model valuation process is a general approach to

FIGURE 2 Interplay between the model and the various valuations.



identifying valuation $v(\pi_{t+h})$. Compared to those methods applied in the banking industry, an additional accounting step is applied.

Note that for both outputs, the forecast and the valuation, models and hence model risk are involved. Bear in mind that forecast and valuation are two different problems. Whereas for the valuation the underlying probability distribution of the model is of interest, in the forecast situation a complete forecast distribution is derived.

Derman (1996) was the first one who referred to model risk in the field of financial engineering, stating that models give at least a rough approximation of reality. Given the relevance of that topic, it is surprising that the number of publications devoted to model risk is rather small.

Rebonato (2001) defined model risk as

> . . . the risk of occurrence at a given point in time (today or in the future) of a significant difference between the mark-to-model value of a complex and/or illiquid instrument held on or off the balance sheet of a financial institution and the price at which the same instrument is revealed to have traded in the market – by brokers' quotes or reliable intelligence of third-party market transactions – after the appropriate provisions have been taken into account.

This definition refers only to a financial instrument but neglects errors made when forecasting the value of the institution's portfolio. Note also that value-at-risk (VaR) models are large-scale models, whereas pricing models are comparatively small.

We assume that model risk stems from the discrepancy between the implemented data-generating process and the data that is at hand or that may be gathered in the future. In the following we focus on a definition that comprises estimation errors and misspecification risks, but we consider implementation risk only insofar as this type of error might contaminate our data.

The quantification of the model risk is also still in its infancy. This problem is most obvious in the context of financial risk management and portfolio valuation, especially of derivative instruments.

This paper aims at developing a working definition of model risk and giving an overview of the current practice of model risk quantification. Section 2 focuses on the main types that the term "model risk" comprises. Sections 3 and 4 review the actual practice of model risk quantification in the market risk measurement and derivatives' valuation context. Section 5 discusses robust methods to avoid model risk due to biased estimates because of data contaminations, and sketches the application of stress tests. Section 6 concludes.

2 TYPES OF MODEL RISK

Although the importance of model risk comes more and more into the focus of practitioners as well as researchers, so far there is no clear notion of how model risk should be defined. Basically there are two different, extreme notions of model risk. A more practice-oriented viewpoint is that everything that might be related in one way or another to the statistical model used is part of the model risk. This includes data contaminations as well as a wrong implementation or a wrong choice of starting values. A wrong application of the statistical model is also seen as a source of model risk. In this construct, even the behavior, preferences and abilities of the statistician are seen as part of the model risk. For an overview of this notion of model risk, see Crouhy *et al* (1998).

Although a wrong implementation of the model is definitely a problem and poses a risk for the financial institution, this should not in our opinion be part of the risk. Therefore, in our setup these sources of risk are treated as part of the operational risk of the financial institution, but not as part of the model risk.

The other line of definition is rather research-oriented and strictly mathematical; see Kerkhof *et al* (2002), who define a statistical model as a probability space. Although this approach allows for a high generality, it does not seem to be handy for practical purposes. The probability space of a model uniquely defines a model and thus contains all the necessary information. However, it is a very abstract approach, not giving insight into the actual statistical modeling procedure. This is mainly because a new parameter space is needed for every parameter constellation and model specification to avoid identification problems. Therefore, the set of valid probability spaces to be evaluated has to be massive, and it is impossible to set it to sensible limits in most practical situations. Another drawback is that it focuses on the distributional properties, whereas in the situation of forecasts, the forecast distribution is not the relevant information, as we will discuss later.

In our approach, model risk can occur at different levels of the statistical modeling procedure. The first level is the model risk induced by the probability space, as defined by Kerkhof *et al* (2002). On a second level, the modeler has to draw conclusions as to which concrete econometric model specification is adequate. This is usually done by the application of various specification tests. However, it is well known that the chosen model can depend on the choice of tests applied to the data. Therefore, the choice of tests is a source of model risk as well. On a third level the modeler has to estimate the model parameters. The wrong choice of an estimator can also induce model risk.

Our approach is on a middle ground between these two extreme approaches. It comes from the belief that the usual procedure for measuring risk is to fit a model to the data at hand, either a portfolio of financial derivatives or a portfolio of credits or anything else, and then draw conclusions about possible losses based on the dynamics of the fitted model. By introducing a model, we at least implicitly give all statistically relevant information, such as probability measures. Model risk here is the risk that this fitted model is wrong. Thus, it does not draw any conclusions from the fitted model but measures the difference between the dynamics of the actual data and the fitted model. This understanding of model risk comes from the econometric modeling approach discussed, for example, in Clements and Hendry (1996).

We first give a rather broad definition of model risk in line with the definition of Crouhy *et al* (1998).

DEFINITION 1 Every risk induced by the application of a statistical model is called model risk.

It should be mentioned that under the notion "statistical model" we subsume statistical, stochastic and econometric models.

A more specific approach following a data-driven econometric modeling idea is as follows.

DEFINITION 2 Every risk induced by the choice, specification and estimation of a statistical model is called model risk in the strict sense.

We thus see model risk to be possibly introduced in one of three modeling steps.

REMARK This definition excludes every risk from "human failure". However, it still includes the risk induced by data contamination, as this might lead to the use of robust procedures.

We assume a model based on a vector of observations x_t including all variables in period t. Given $X_t = (x_{t-1}, \dots, x_1)$ the joint probability of x_t can be stated as:

$$\prod L(x_t \mid X_{t-1}; \Theta)$$

This is in line with Kerkhof *et al* (2002), who specify the model by a probability space. The concrete specification of the model contains four steps, including the estimation and specification steps from Definition 2:

1) marginalization of the data generating process;
2) model specification with respect to the choice of variables;
3) model specification with respect to the functional form; and
4) estimation of the parameters.

For further details see Cuthbertson *et al* (1992). Errors may occur for any of the following reasons:

TABLE 1 This table summarizes major components of model risk.

Component of model risk	Category	Procedure step #
Model errors and misspecification	• Errors in the analytical solution	3
	• Misspecification of the underlying stochastic process	1
	• Missing risk factors	2
	• Missing considerations (eg, transaction costs)	2
	• Misclassifying or misidentifying the underlying asset	2, 3
	• Changing market conditions	1, 2, 3
	• Length of sampling period	1
	• Errors in variables	2
Model estimation	• Competing statistical estimation techniques	4
	• Estimation errors	4
	• Outlier problems	2, 4
	• Estimation intervals	4
	• Calibration and revision of estimated parameters	4

We would like to emphasize that our approach follows classical statistical reasoning. Others, eg, Crouhy *et al* (1998), consider the whole process in a more flexible fashion by allowing the interaction of the statistician and the statistical model. Given the complexity of a VaR model, we see no good way to carry out a valid modeling. This might be sensible only for pricing models.

- parameter change;
- model misspecification;
- estimation uncertainty;
- variable mismeasurement;
- initial condition uncertainty;
- non-modeled variable uncertainty;
- incorrect categorization of some variables as exogenous;
- lack of invariance to policy changes in exogenous variables; or
- error accumulation.

This approach bridges the gap between the abstract notion of Kerkhof *et al* (2002) and the more practical statistical modeling approach.

As outlined in Table 1, our data-driven modeling emphasizes two central tasks: choosing the functional form, and using the data to estimate the parameters. Model mismatch errors mirror the model's inability to represent the data. Estimation errors end up in incorrect parameter values. The well-known bias-variance trade-off simultaneously hinders the reduction of errors from both underfitting and overfitting. Therefore, we assume that the principle of parsimonious modeling is implicitly fulfilled. Note that overfitting will decrease the in-sample error; however, the out-of-sample error will in general start to increase. The latter type of error is of particular importance for forecast distributions. Considering a forecast problem, the model error is of minor importance, because forecast quality is judged by a function of the forecast distribution and the realization:

$$T(L(X_{t+h} \mid I_t), \nu(\pi_{t+h}))$$

Of course, model errors might deteriorate forecast quality. Note that in Kerkhof *et al* (2002) the model error is measured by a metric that compares different forecast distributions. We follow approaches summarized in Corradi and Swanson (2006).

Like others, our approach is not free of problems. To measure the risk induced by a model, it has to be compared to an alternative model. To correctly specify the model risk you have to know an accurate model. However, accurate models are often hard to find. As we cannot always find an accurate model we need a benchmark model giving a close approximation to our data. The problem is how to choose this benchmark model. If we knew the model closest to our data, we could easily use this model for risk measurement, and model risk would no longer be a problem. Thus, finding a suitable benchmark model is a serious problem. However, it will hardly be possible to overcome this problem when discussing model risk.

In this respect the problem of quantifying model risk is closely connected to concepts of robust statistics, where the influence of wrong data on an estimator or test is considered. These concepts are directly applicable in our setup in the second and especially in the third step of the modeling procedure. If the data is contaminated the use of a robust estimator might be preferable, as is discussed in Section 5.

Statistical models in finance are basically used for two major applications: firstly they are used for the pricing of derivatives and secondly they are used to measure the market or credit risk of a financial institution. These two applications lead to different problems when measuring the model risk. Whereas in the first case the model has to be evaluated and thus slight departures of the model can lead to huge errors, in the second case the model is used for forecasting. It should be mentioned that forecasts can be quite robust against model misspecifications. For a detailed discussion and an example, see Hamerle and Rösch (2005).

3 MODEL RISK WHEN MEASURING RISK

The development of sophisticated credit risk models is still a challenge to the risk management industry. Specifically the issue of non-linear dependency of credit risks poses big problems. When we use different copulas or different correlations, profit/loss distributions can deviate widely. The inherent model risk can hardly be quantified. The same problem holds for operational risks. Models for the quantification of market risk, however, are much more developed. Therefore, we concentrate on the model risk measurement for market risks. Portfolio risk can then be decomposed into market risk, estimation risk and misspecification risk.

Currently, the Basel Committee sets a multiplication factor of three in order to take model risk into account (see Stahl (1997)). This approach is well justified, as the performance of VaR models through the 2001 crisis has shown (see Jaschke *et al* (2007)). According to the coherence theory of Artzner *et al* (1999), the calculated model risk measure should be added to the market risk measure to make a risky position acceptable.

There are two models used in the context of measuring the model risk inherent in market risk models: the Bayesian approach and the worst-case approach.

3.1 The Bayesian approach

Suppose that K comprises all candidate models to measure market risk. Then the following prior beliefs are necessary: the priors on the model parameters θ and the prior beliefs of the probability P for all models K that this model is the true model. Then the Bayesian approach calculates the expected risk measure by taking the average over all candidate models K. In the context of a model risk measure for market risk, this means calculating VaR with different models, eg, historical simulation, stochastic simulation and variance–covariance approach, and then calculating the average of the resulting VaR.

The strengths of this approach lie in its predictive ability and the stability of the resulting estimates. The main shortcomings of this approach have to do with the difficulties of measuring the priors and the probability of each model.

3.2 The worst-case approach

Let a model m be given in a class of models M. A financial asset's risk on the model m is denoted by π_m. K is the tolerance set of all models that describe the alternative dynamics including the nominal model m. Model risk under the worst-case approach is defined as the difference between the market risk measure of a nominal model and the market risk measure under the worst-case model, formally:

$$\rho(\pi, m, K) = \sup_{k \in K} RM_k(\pi_k) - RM_m(\pi_m)$$

(see Kerkhof *et al* (2002)), where $\sup_{k \in K} RM_k(\pi_k)$ is the worst-case risk measure. Suppose that model risk is the sum of estimation risk and misspecification risk as defined in Section 2. Estimation risk is the risk of choosing a wrong element $m(\hat{\theta})$ out of a model class $M(\theta)$ even if the actual dynamics belong to it. One typically proceeds by defining a confidence region around the estimator $\hat{\theta}$. The risk of wrong estimation is also defined as model risk restricted to the model class, whereas misspecification risk is unrestricted to $M(\theta)$. The alternative dynamics can then be distinguished by a restricted tolerance set K_r and an unrestricted tolerance set K_u.

This term can be non-negative. To overcome this weakness, the restricted tolerance set has to be nested within the unrestricted tolerance set, for example by forming convex combinations of models or by using a family of unrestricted tolerance sets that have been parametrized by a confidence level. Let α and K_r be given. Then $K_u = K_u(\beta)$ with $\beta = \min(\alpha, \sup \gamma \in (0, 1) : K_r \in K_u(\gamma))$ (see Kerkhof *et al* (2002)). This approach fails in accurately quantifying the risk that stems from using a wrong model, because the actual dynamics are still unknown. The approach equates the worst-case scenario risk measure with the risk under the worst case.

These considerations refer only to the first step of our modeling procedure. Kerkhof *et al* (2002) focus on the comparison of forecast distributions when measuring model risk. Following the model-building steps two to four in the previous section can reduce the model risk significantly by correctly specifying and estimating the model. Although each of these steps can induce an error itself, this modeling procedure will result in a model that is as adequate as possible, in the sense that it gives a reliable forecast. Thus, our model-building approach focuses on the comparison of forecast realization of the underlying data-generating process, or an adequate approximation of it and the specified model, and therefore it goes beyond the Kerkhof approach.

4 MODEL RISK WHEN PRICING DERIVATIVES

An important component of a financial institution's model risk results from the valuation of derivative instruments. In the context of option pricing, the importance of this topic becomes obvious when we allow for the fact that a bank usually steps in to the (riskier) writer side of the option. The liability of the buyer is limited to the premium paid for obtaining the right to execute the option, so that private investors prefer this side of the contract (see Green and Figlewski (1999)).

A huge number of pricing models proposed in the aftermath of the Black–Scholes–Merton model take into account stochastic volatility and the empirically established fact that the returns of the underlying asset exhibit excess kurtosis; that is, extreme events are more likely than under the assumption of lognormal returns. Nonetheless, the Black–Scholes model is the most widely applied model for valuing options (see Bakshi *et al* (1997)). Besides misspecifying the pricing rule, another crucial model risk source is related to the estimation of the required input parameters to value a derivative, especially the volatility that has to be forecasted. The modeler can choose between estimating the volatility from historical data, adopting conditional volatility models from the (generalized) autoregressive conditional heteroskedasticity ((G)ARCH) family or using the model-implied volatility.

A firm could limit its derivative risk by adopting coping strategies such as diversification, cashflow matching or delta hedging (see Green and Figlewski (1999)). The last of these is the most common method.

To quantify the model risk induced by derivatives valuation, the just-described approaches can be applied. The first method involves averaging the option prices obtained from applying different pricing methods in the Bayesian framework. Another possibility would be to differentiate between market and model risk under the worst-case approach.

Let S be the underlying risk of the option, $(H_i)_{i\in I}$ be its payout and $(C_i^*)_{i\in I}$ be the current price observed in the market, which ranges within a span $(C_i^*) \in [C_i^{\text{bid}} - C_i^{\text{ask}}]$. Denote model risk by a mapping $\mu : C \mapsto [0, \infty[$ and by $\mathbb{Q}$ a pricing rule whose function is to relate the price of the underlying asset to the price of the derivative instrument under the assumption of no-arbitrage, ie, the martingale representation should be fulfilled. λ stands for the fraction of the spread attributable to model risk. Denote by:

$$C = \left\{ H \in F_T,\ \sup_{\mathbb{Q}\in Q} E^{\mathbb{Q}}[|H_i|] < \infty \right\}$$

a contingent claim for each model $\mathbb{Q}$. Arbitrage freedom is secured by the martingale $G_t(\phi) = \int_0^t \phi_u \,\mathrm{d}S_t$, where $(\phi_t)_{t\in[0,t]}$ is a predictable process. Cont (2004) assumes that the price of a liquid plain-vanilla option forms itself within the interaction of supply and demand and contains no model risk. Hence, model uncertainty in the derivatives context is a problem only for over-the-counter (OTC) instruments whose price cannot be determined easily. Instead, a pricing rule needs to be applied to value an OTC instrument.

The upper price bound of the contingent claim of the firm is defined as:

$$\bar{\pi} = \sup_{\mathbb{Q} \in Q} E^{\mathbb{Q}}[X]$$

and the lower price bound as:

$$\pi = \inf_{\mathbb{Q} \in Q} E^{\mathbb{Q}}[X]$$

The result of applying a pricing rule will be a risk measure within the range spanned by the lower and upper bounds. A model risk measure fulfilling the axioms above is then defined by:

$$\mu_Q(X) = \bar{\pi}(X) - \underline{\pi}(X)$$

Hence, there will be no risk if the payout is not influenced by model uncertainty:

$$\bar{\pi}(X) = \underline{\pi}(X)$$

From these considerations it becomes clear that the model risk in the context of option pricing is measured by the comparison of distributions. This is in contrast to the type of model risk used for risk measurement. Bear in mind that the distributions compared are specified by the model resulting from our model-specification procedure detailed in Section 2. Therefore, our framework covers this type of model risk as well.

Cont (2004) assumed that liquid options do not include model risk. This assumption is quite surprising, as by assessing a derivative price using a valuation formula, market-makers include model risk when calculating the option price.

5 MODEL RISK INDUCED BY CONTAMINATED DATA

As described in Section 2, there are different levels where model risk may occur. One important problem is reliable parameter estimation. One common reason for wrong parameter estimates is contaminated data. This can be due to human failure or the simple impossibility of obtaining uncontaminated data such as for generating forward projections. Whatever the reasons may be, it is well known that data contamination can lead to serious biases in the most commonly used estimators and therefore influence the model risk. The following example sheds some light on this phenomenon.

EXAMPLE 1 Estimation of corporate ratings.

In the following we consider a logistic model underpinning an internal rating system. The input data B_0 consists of balance sheet data for some 1,000 corporations over multiple years, which were coded such that:

$$x_{ij} \in [0, 100] \equiv I \tag{1}$$

where higher values correspond to higher risk. The associated variables for the realizations in B_0 are denoted by:

$$D, X_1, \dots, X_6$$

The probability π that a firm with characteristics $x(t)$ defaults over time period t to $t+1$ is:

$$\pi = P(D(t+1) = 1 \mid X(t) = x(t))$$

which can be estimated via a logistic regression with parameter w (d being the response variable and x^* being the regressors):

$$\pi = P(x, w) = \frac{e^{(w \cdot x*^T)}}{1 + e^{(w \cdot x*^T)}} \tag{2}$$

where:

$$x^* = (1, x_1, \dots, x_6), \quad w = (w_0, w_1, \dots, w_6)$$

and T denotes the transpose, as usual.

Miskeyed values are one of the problems a modeler faces. According to (1) the range of values of the characteristics x_{ij} is a fixed interval. For the data set at hand, however, 5% of the observations violate the condition imposed by (1) and approximately 4.5% of them are defaulted, ie, $d_i = 1$. Note that our analyses focus on those observations that are known *a priori* to be erroneous. Some additional uncertainty might be introduced by those erroneous observations initially remaining undetected.

In the following we consider three modifications of B_0 – denoted by B_1, B_2 and B_3 – that fulfill (1):

$$B_1 = \{x \mid x \in B_0 : x_{ij} \in I\}$$

$$B_2 = \{\tilde{x} \mid x \in B_0 : \tilde{x}_{ij} = x_{ij} \cdot 1_I(x_{ij}) + 100 \cdot 1_{\{x_{ij} > 100\}}(x_{ij}) + 0 \cdot 1_{\{x_{ij} < 0\}}(x_{ij})\}$$

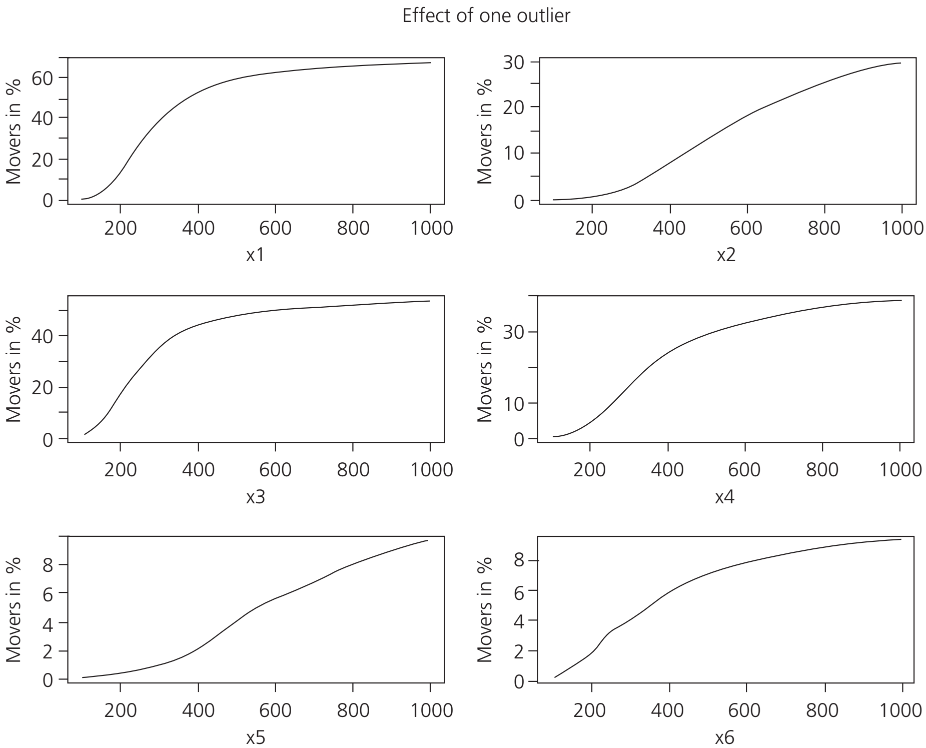
$$B_3 = \{\tilde{x} \mid x \in B_0 : \tilde{x}_{ij} = x_{ij} \cdot 1_I(x_{ij}) + x_{m_j} \cdot 1_{\bar{I}}(x_{ij})\}$$

where 1 denotes the indicator function and x_{m_j} denotes the median of characteristic x_j. Note $B_1 \subset B_0$, so the deletion approach loses some information. The mapping of outliers on the interval bounds in B_2 is the modification that was actually chosen by the bank. Approach B_3 is motivated by simple imputation techniques. It substitutes the median for false values.

As shown in Stahl *et al* (2008), the impact of the different imputation techniques on the estimated $\hat{w}$ is statistically not significant. However, their impact on the parameter of interest, the rating of the corporation, might be of practical relevance, as Table 2 shows.

The detailed analysis in Stahl *et al* (2008) raised the question of what impact an outlier might have on the misclassification of corporates. It is based on a subsample of size 999 from B_1 – where all wrong values were deleted – and one additional outlier was added, taking the values 200, 300, ..., 1,000. Then the whole estimation process was repeated for every covariate. Figure 3 shows that the covariates differ in respect to the sensitivity of the outlier. In any case the impact is highly relevant.

FIGURE 3 For the data set B_1, the effect of one additional non-defaulted observation with extreme values in one of the characteristics is shown.



The six graphics correspond to the six characteristics of the rating model. Each graphic shows the effect of an extreme outlier in the respective characteristic on obligor migration, taking B_1 as a point of reference. The value of the outlier ranges from 100 to 1,000. The interpolation of these points yields the graph.

TABLE 2 This table contains summary statistics of rating movements induced by choices of $B_{(\cdot)}$.

B_0	B_1	B_3	Stress	
$\hat{\mu}^{(k)}$	0.35	0.39	0.42	0.9985
$\hat{\sigma}^{(k)}$	0.51	0.51	0.64	0.04
Percentage of movers	33	38	38	99.85
Maximum move length	6	4	5	1

The mean of the first row indicates an average move size of 0.39 rating grades, with the third row showing that an average of 36% movers are to be expected. The second row yields a standard deviation of slightly more than half a rating notch. Comparing the first row with the third shows that a move of more than one notch is rare. The extremes are given in the last row. The last column shows the summaries for a stress test that grades all obligors down by one notch except those already belonging to the worst non-default rating grade and the defaulted obligors.

One way of overcoming this problem is the use of robust methods. In this section we give a brief review of the most common robustness concepts and estimators suitable for overcoming model risk induced by data contaminations.

There are two different measures of robustness: local and global. Local robustness measures the effect of an infinitesimal deviation from the underlying distribution at any point to the applied estimator. Global robustness, on the other hand, deals with how many data points in the sample can be contaminated without the estimator's breaking down.

Local robustness is measured in terms of the influence function of an estimator. In order to define the influence function, let F denote a parametric family of distributions depending on the parameter θ to be estimated. Let $T(x_1, \dots, x_n)$ denote an estimator depending on a sample $x_1, \dots, x_n$. Furthermore, denote by Δ_x the one-point mass at x. The influence function of T at the point x is defined by:

$$IF(x; T, F) = \lim_{t \downarrow 0} \frac{T((1-t)F + t\Delta_x) - T(F)}{t}$$

Thus, the influence function measures the sensitivity of the estimator to infinitesimal deviations from the true distribution. A related robustness concept using the influence function is B-robustness, where the B stands for bias. A measure for B-robustness is the gross-error sensitivity defined by:

$$\gamma^* = \sup_x |IF(x; T, F)|$$

By calculating the gross-error sensitivity, we have a measure of the worst possible effect that a small contamination of a fixed size can have on the estimator. That it is an upper bound of the asymptotic bias of the estimator illustrates where the name of the associated robustness concept comes from. For a more detailed overview of the influence function and related robustness concepts, see Hampel *et al* (1986).

When considering local robustness it is also of interest to consider global robustness. Global robustness is measured by the breakdown point: the fraction of observations that can be contaminated without affecting the estimator. Obviously, an optimal value to be achieved by a robust estimator is a breakdown point of 1/2. For the formal definition of the breakdown point, again let T be an estimator for the sample $(x_1, \dots, x_n)$. It is defined by:

$$\varepsilon^*(T, x_1, \dots, x_n) = \frac{1}{n} \max\Big\{m; \max_{i_1,\dots,i_m} \sup_{y_1,\dots,y_m} |T(z_1, \dots, z_n)| < \infty\Big\}$$

where the sample $(z_1, \dots, z_n)$ is obtained by replacing the m data points $x_{i_1}, \dots, x_{i_m}$ by the arbitrary values $y_1, \dots, y_m$. The above version of the breakdown point is also often called a finite-sample breakdown point to distinguish it from an asymptotic version, which is not considered here.

Bear in mind that this definition of the breakdown point covers only "explosion" of the estimator, which might not be the only problem for an estimator. When estimating the variance, "implosion" to zero can also be a problem. This can easily be covered by including the minimum and infimum in the definition as well. We will omit this here for the sake of notational simplicity. The version given here is especially designed for location estimators.

As model risk can be induced by local as well as by global robustness, it is desirable to have an estimator that covers both problems. One possibility is the use of M-estimators. These by definition satisfy the equality:

$$\sum_{i=1}^{n} \psi(X_i, T) = 0$$

where ψ is any suitably chosen function. In order to guarantee robustness in any sense, the function should be chosen to be redescending. A possible choice of ψ guaranteeing B-robustness and a breakdown point of $1/2$ is Tukey's biweight:

$$\psi(x) = x(r^2 - x^2)^2 1_{[-r,r]}(x)$$

for some value of r.

M-estimators have the undesirable property of not being scale-invariant. Therefore, Rousseeuw and Yohai (1984) introduced S-estimators based on minimization of a scale statistic rather than a location statistic. To define S-estimators, define the scale statistic $s(r_1, \ldots, r_n)$ by:

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{r_i}{s}\right) = K$$

with some constant K and the function ρ to be chosen, for instance so that its derivative is Tukey's biweight. Then, the S-estimator of location $\hat{\theta}$ is defined by minimizing $s(x_1 - \hat{\theta}, \ldots, x_n - \hat{\theta})$ over all $\hat{\theta}$. This simultaneously gives the scale estimate:

$$\hat{\sigma} = s(x_1 - \hat{\theta}, \ldots, x_n - \hat{\theta})$$

The influence function of an S-estimator is equal to that of the location M-estimator; its breakdown point is $1/2$ as well. S-estimators were especially constructed for regression models and multivariate situations where they have good robustness properties. They are also well-understood in many other situations of practical interest. For instance, Sibbertsen (2001) derived the properties of S-estimators under long-range dependencies and trends, showing their applicability in various situations. As these are situations figuring in many econometric models for measuring risk, these estimators seem to be a suitable choice to help avoid model risk due to data contaminations.

Our exposition showed that a robust model merits consideration. Robust methods are useful when some outliers contaminate the data but the overall model assumptions can still be assumed to be valid. In times of crisis this is not always the case, as events are too extreme for any model to hold and the assumed model breaks down completely. Therefore, modern risk management processes do not and cannot stop with the implementation of models devoted to measuring risks. In order to capture these aspects of model risk, so-called stress tests are run in practice. These are an integral part of risk management processes, especially for those adverse developments where economic capital – the most important parameter derived from a VaR model – is of no help. The current liquidity crisis is a good example. Typically, stress tests reflect scenario-based calculations, where the scenarios represent extreme situations. Our robust framework is to a large extent immune against these extremes. Therefore, we propose to apply stress tests not for the determination of the forecast distribution, $L(X_{t+h} \mid I_t)$, but for the evaluation of the portfolio at hand.

One approach that would fit smoothly with our robust framework is the application of the configural sampling approach introduced by Morgenthaler and Tukey (1991). These authors suggest, for example, mixtures of Cauchy distributions and other extremes in the set of distribution functions in order to challenge estimators. This approach generalizes scenario-based stress tests to distribution-based ones. However, this approach misses one of the major challenges in deriving stress tests: the requirement that they refer to extreme, although still plausible, situations. To that end recent contributions from Balkema and Embrechts (2007) seem promising. They define high-risk scenarios as follows.

DEFINITION 3 Given a random vector Z in R^d, for any closed halfspace H in R^d with $P(Z \in H) > 0$, the high-risk scenario Z^H is defined as the vector Z conditioned to lie in the halfspace H. If Z has distribution μ then Z^H has the high-risk distribution μ^H, where:

$$\mathrm{d}\mu^H(Z) = 1_H(z)\,\mathrm{d}\mu(z)/\mu(H)$$

It is shown that the multivariate generalized Pareto distribution is the limit distribution for high-risk distributions. In Bonti *et al* (2007) a conditional approach that is in line with the spirit of this definition was implemented; however, further practical experiences have to be gathered in order to implement the new approach outlined in Definition 3.

6 CONCLUSION

Because of the steadily more frequent use of sophisticated statistical models for measuring risk or pricing derivatives in financial institutions, the problem of measuring the risk induced by the usage of a wrong model becomes more important. Even though practitioners as well as researchers are aware of this worsening problem, there is hardly any literature on the subject, let alone a unified approach and notion of what model risk actually is. In this paper, therefore, we provide a review of the existing, rather extreme, approaches and develop a compromising notion of model risk, driven by a data-oriented model-building procedure. Our approach covers model risk when measuring the risk of a financial institution as well as model risk when pricing derivatives. In this respect our approach even extends the previous notions, as they do not distinguish between these two types of model risk and lack practical applicability when it comes to forecasting. In order to minimize the risk induced by our model-building strategy, we further propose the use of robust estimates when the data might be contaminated.

REFERENCES

Alexander, C. (2003). The present and future of financial risk management. *Journal of Financial Econometrics* **3**, 3–25.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance* **9**, 203–228.

Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance* **52**, 2003–2049.

Balkema, G., and Embrechts, P. (2007). *High Risk Scenarios and Extremes*. European Mathematical Society, Zurich.

Baumgartner, G., Bühlmann H., and Koller, M. (2004). *Multidimensional Valuation of Life Insurance Policies and Fair Value*. Mitteilungen SAV I, Schweizerische Aktuarvereinigung, Bern.

Bonti, G., Kalkbrener, M., Lotz, C., and Stahl, G. (2007). Credit risk concentrations under stress. *The Journal of Credit Risk* **2**(3), 115–136.

Clements, M. P., and Hendry, D. F. (1996). Forecasting in macro-economics. *Time Series Models.* In *Econometrics, Finance and Other Fields*, Cox, D. R., Hinkley, D. V. and Barndorff-Nielsen, O. E. (eds). Chapman & Hall, London, pp. 101–141.

Cont, R. (2004). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance* **16**, 519–547.

Cont, R., Deguest, R., and Scandolo, G. (2007). Robustness and sensitivity analysis of risk measurement procedures. Financial Engineering Report No. 2007-06, Columbia University Center for Financial Engineering.

Corradi, V., and Swanson, N. R. (2006). Predictive density evaluation. *Handbook of Economic Forecasting*, Elliot, G., Granger, C. and Timmermann, A. (eds). Elsevier, Amsterdam, pp. 197–284.

Crouhy, M., Galai, D., and Mark, R. (1998). Model risk. *Journal of Financial Engineering* **7**, 267–288.

Cuthbertson, K., Hall, S. G., and Taylor, M. P. (1992). *Applied Econometric Techniques*. Harvester Wheatsheaf, York, UK.

Derman, E. (1996). Model risk. *Quantitative Strategies Research Notes*, Goldman Sachs.

Green, T. C., and Figlewski, S. (1999). Market risk and model risk for a financial institution writing options. *Journal of Finance* **54**, 1465–1499.

Hamerle, A., and Rösch, D. (2005). Misspecified copulas in credit risk models: how good is Gaussian? *The Journal of Risk* **8**, 41–58.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, Inc, New York.

Jaschke, S., Stahl, G., and Stehle, R. (2007). Value-at-Risk forecasts under scrutiny – the German experience. *Quantitative Finance* **7**, 621–636.

Kerkhof, J., Melenberg, B., and Schumacher, H. (2002). Model risk and regulatory capital. Available at SSRN: http://ssrn.com/abstract=301531.

Marshall, C. (2000). *Measuring and Managing Operational Risks in Financial Institutions*. Wiley, New York.

Morgenthaler, S., and Tukey, J. W. (1991). *Configural Polysampling*. Wiley, New York.

Padoa-Schioppa, T. (2004). *Regulating Finance*. Oxford University Press.

Rebonato, R. (2001). Managing model risk. *Mastering Risk*, Alexander, C. (ed). Prentice Hall Financial Times Press, London, Vol 2, pp. 82–115.

Rousseeuw, P., and Yohai, V. J. (1984). Robust regression by means of S-estimators. *Robust and Nonlinear Time Series Regression (Lecture Notes in Statistics, Vol. 26)*. Springer, New York, pp. 256–272.

Sibbertsen, P. (2001). S-estimation in the linear regression model with long-memory error terms under trend. *Journal of Time Series Analysis* **22**, 353–363.

Stahl, G. (1997). Three cheers. *Risk Magazine* **10**, 67–69.

Stahl, G., Nill, E., Siehl, B., and Willsberg, J. (2008). Risk modeling and model risk – the IRBA case. Working Paper, Bundesanstalt für finanzdienstleistungsaufsicht.

UniCredit Group (ed) (2008). Insurance: risk-adjusted insurance-company valuation.