Modelling credit risk in the enterprise sector – further development of the SEBRA model

Eivind Bernhardsen, adviser, and Kai Larsen, special adviser, Financial Markets Department, Norges Bank[1]

Since 2001, Norges Bank has used an empirical model, the SEBRA model[2], to estimate bankruptcy probabilities for Norwegian limited companies. The model is also used to estimate banks' expected losses on loans to enterprises in different industries. This article presents two new versions of the model: an extended version of the original model, and a basic version which makes less use of variables which correlate with the size of the enterprise. We show that the basic version is better suited to predicting and projecting banks' overall loan losses. However, the accuracy rate for bankruptcies is slightly lower at enterprise level. The extended version is better suited to analyses where the emphasis is more on bankruptcies than on aggregate loan losses.

1. Introduction

Norges Bank's SEBRA model estimates bankruptcy probabilities using key figures calculated on the basis of enterprises' annual accounts, and information on their age, size and industry classification. Multiplying these bankruptcy probabilities by each enterprise's bank debt and then adding up the figures for all enterprises gives us an estimate of banks' expected loan losses due to bankruptcy, assuming that the entire loan amount is lost. Analyses based on such estimates are published regularly in Norges Bank's report *Financial Stability* and are included in its continuous assessment of the outlook for banks' financial strength. In analyses of enterprises' credit risk, we look at the situation both in different industries and in different regions. The SEBRA model is also used for projecting and stress testing banks' loan losses in various macro scenarios, for analyses of banks' pricing of loans to enterprises, and for assessing the potential effects of changes in the capital adequacy rules.[3] Kredittilsynet (the Financial Supervisory Authority of Norway) uses bankruptcy probabilities from the model in its on-site supervision of banks and in its analyses of the state of financial markets.

This broad use of the SEBRA model has over time provided useful experience and ideas for further development over the years. In addition, access to data has improved since the model was developed. The original SEBRA model's accuracy rate for bankruptcy at enterprise level has been high and stable over time. The model also captures the surge in banks' recorded loan losses during the banking crisis of the early 1990s. However, the next increase in banks' loan losses, which came in 2002 and 2003, is not captured to the same extent.

In this article, we look more closely at various needs for the further development of the SEBRA model. We present two new versions of the model: an extended version of the original model, and a basic version which uses a smaller number of explanatory variables. After evaluating the accuracy and predictive power of these models, we describe briefly how banks' recorded loan losses can be projected. The article concludes with a summary.

2. The original SEBRA model in brief

In the original SEBRA model, the probability of bankruptcy is modelled mainly using key figures for an enterprise's earnings, financial strength and liquidity, see Eklund et al. (2001). Thus, the model's predictions are driven by quantities that reflect key business economic conditions at the individual enterprise. These will always be crucial for an enterprise's capacity to service its debt. Besides key financial figures, the model includes measures of an enterprise's size and age, and industry variables based on aggregates of the key financial figures. It is useful to differentiate between variables which reflect financial conditions and variables which are more indirectly related to these conditions but still contribute to the model's overall explanatory power. Examples of the latter are the level of tax payable, trade accounts payable and dividend provisions.

The model does not include additional information such as negative credit history, absence of auditor approval, or late or non-filing of annual accounts. This ensures that the model attaches more importance to the financial factors behind movements in risk, which is important given that the model's main purpose is to contribute to an understanding of movements in credit risk. Furthermore, it would be very difficult to project such variables. The model is also more stable, as experience shows that the registration quality of this additional information varies from year to year. The model does not take explicit account of historical variations in bank-

[1] We would like to thank Kåre Hagelund, Kjell Bjørn Nordal, Snorre Evjen, Arild Lund, Bjørn Helge Vatne and Bjørne Dyre Syversten for useful comments and contributions.
[2] The acronym SEBRA derives from the Norwegian for "System for EDP-based Accounts Analysis".
[3] See, for example, Frøyland and Larsen (2002), Bernhardsen and Larsen (2002), and Larsen and Bjerkeland (2005).

ruptcy frequency between industries. These differences are instead represented through variables for industry averages and variances of basic key variables based on a detailed industry classification. In this way, changes in risk levels in different industries over time can be captured, and the model becomes less retrospective.

3. The need to further develop the SEBRA model

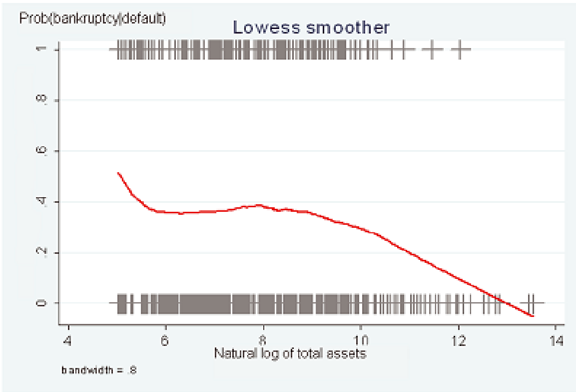
Long experience of the use of the SEBRA model has meant that we have discovered various weaknesses in it. In this section, we discuss the most important needs for improvement. There are also other reasons to reassess the model. For example, the way in which the explanatory variables are measured in enterprises' annual accounts may have evolved over time, due in part to new accounting rules. There may also have been changes in the registration of bankruptcies over time. Access to new and more data is another factor which makes the further development of the model desirable.

Better prediction of the risk of losses on loans to large enterprises

The risk of losses is not the same as the risk of bankruptcy. The original SEBRA model's accuracy rate for bankruptcy at enterprise level has generally been high and stable over time. In the original SEBRA model, size (measured as the logarithm of total assets) is included as an explanatory variable. It appears that small enterprises go bankrupt more often than large enterprises for given values of the explanatory variables. If this size effect applies less to the probability of a loan loss, it will be problematic using bankruptcy as a substitution variable for losses in a model that uses size as an explanatory factor. Such a model will overestimate the effect of size on defaults and losses. Small enterprises often have little bank debt in NOK. In many cases, therefore, it is the tax authorities or suppliers who file bankruptcy petitions for these enterprises. In the event of problems with larger loan exposures, however, banks often play an active negotiating role. This may result in all or parts of the exposure being recognised as a loss while the enterprise avoids bankruptcy petition and proceedings.

Defaults are probably a better indicator of losses than bankruptcies. We have information on defaults for only a limited sample of enterprises and cannot, therefore, use defaults to estimate the model. This sample can, however, be used to investigate our hypothesis concerning the size of an enterprise. The grey bars in Chart 1 show non-bankruptcy observations (0) and registered bankruptcies (1) for enterprises of different sizes, all of which have had their loans classified as in default. The red curve is an estimate of the probability of bankruptcy given default. We see that the probability of bankruptcy given default is stable at around 40 per cent for enterprises with total assets below NOK 10 million.[4] After this, the probability begins to fall significantly.

Chart 1 Probability of bankruptcy given default and firm size



Source: Norges Bank

The original SEBRA model includes explanatory variables which are either directly or indirectly related to an enterprise's size. This means that an enterprise with weak earnings and financial strength will nevertheless be assigned a low bankruptcy probability if it is sufficiently large. According to our hypothesis about the importance of size, which is supported by Chart 1, the actual risk of loss may be considerably higher for such enterprises. Systematic underestimation of the risk of loss on loans to large enterprises is particularly problematic in analyses of financial stability, as large enterprises are heavily weighted when calculating expected loan losses. Since the model is non-linear, underestimation of this risk will lead to underestimation of all the explanatory variables in the model.

By developing a model which attaches less weight to variables related to an enterprise's size, the underestimation of the risk of loss associated with large enterprises can be limited. Examples of variables in the SEBRA model which are directly or indirectly related to an enterprise's size include total assets, trade accounts payable, and government taxes payable relative to total assets.

The original SEBRA model largely captures the surge in banks' recorded loan losses during the banking crisis of the early 1990s. However, the next increase in banks' loan losses, which came in 2002 and 2003, is not captured to the same extent. In these years, there was a temporary dip in the competitiveness of many large exporters. Smaller – and often sheltered – enterprises are more dependent on domestic purchasing power, which deteriorated only slightly. Underestimation of the risk of loss on loans to large enterprises may have contributed to the increase in banks' overall loan losses being captured by the model only to a limited extent during this period. In Section 4 below, we present a simplified version of the SEBRA model which attempts to take account of these factors.

[4] In other words, the logarithm of total assets (measured in thousands of NOK) is less than approx. 9 in Chart 1.

Avoiding undesirable effects of changes in dividend taxation

The original SEBRA model includes an indicator variable for dividend provisions to capture expectations of future earnings. Dividend taxation has changed over time. In 2006, for example, tax was introduced on dividends to private shareholders over a stipulated risk-free deduction. This change was announced several years in advance and probably impacted on enterprises' dividend provisions ever since their 2003 accounts.[5] When dividends reflect adaptations to tax changes rather than earnings expectations, the dividend variable will make undesirable contributions to the model estimates. We have not, therefore, included the dividend variable in the two new variants of the model.

Improved system for projections and stress tests

Projections and stress tests of banks' loan losses are becoming increasingly important in analyses of financial stability. In recent years, Norges Bank has used an accounts-based projection and stress testing method together with the SEBRA model. This method was used, for example, in the work on the IMF's stress testing of the Norwegian financial sector, see Hagen et al. (2005), and for stress test analyses in Norges Bank's report *Financial Stability*.

One important challenge in analyses of this kind is to find a good way of projecting key figures. A model which includes large numbers of explanatory variables is more difficult to project than a model with few variables. It is also easier to project basic key figures for the risk drivers earnings, financial strength and liquidity than variables which reflect these drivers more indirectly. Furthermore, it will be easier to explain what is happening in the projections. The need for a more suitable projection method is an important reason why we have chosen to develop a simplified version of the original SEBRA model.

4. Two new versions of the SEBRA model

We have developed two new versions of the SEBRA model: SEBRA Basic and SEBRA Extended (see Table 1). The table shows which explanatory variables are included in the two models.

SEBRA Basic

The basic version includes the original basic key figures for earnings, financial strength and liquidity. Like the original model, it also includes the enterprise's age and a modified indicator variable for impaired equity.[6] We have also introduced a set of industry variables based on basic key figures for earnings and financial strength which vary more over time than in the original model. Previously the industry variables were calculated for the entire estimation period. We now calculate most of the industry variables on an annual basis.

The bankruptcy probabilities for large enterprises produced by SEBRA Basic are consistently higher than with the original SEBRA model and SEBRA Extended. This is primarily a result of SEBRA Basic including fewer size-related variables, which – other things being equal – serve to reduce the bankruptcy probability for large enterprises (see discussion above). However, the average bankruptcy probability is the same in the various versions of the model.[7]

Table 1. Variables included in SEBRA Basic (darker shaded areas) and SEBRA Extended (entire table)

Variable definition	Variable type	Varies by
Ordinary profit before depreciation and write-downs as a percentage of total debt	Key figure	Enterprise/year
	Average	Industry/year
	Standard deviation	Industry/year
	Correlation with Norway portfolio	Industry
Equity as a percentage of total assets	Key figure	Enterprise/year
	Average	Industry/year
Book equity less than paid-in equity	Indicator	Enterprise/year
Liquid assets less short-term debt as a percentage of operating revenues	Key figure	Enterprise/year
Age (years) = 1, 2, 3 … 8	Indicators	Enterprise/year
Total assets in fixed NOK	Key figure	Enterprise/year
Trade accounts payable as a percentage of assets	Key figure	Enterprise/year
Unpaid taxes and dues as a percentage of assets	Key figure	Enterprise/year

[5] Dividends set aside in the accounts for year t are paid and taxed in year t+1.

[6] When calculating this variable, we adjust paid-in equity for historical write-downs. This is done to counteract the effects of enterprises' adjustment to the introduction of tax on dividends to private shareholders on 1 January 2006. Provided that various criteria are met, shareholders can still take out dividends tax-free by writing down paid-in equity.

[7] In the logit model, the average predicted bankruptcy probability will always coincide with the overall bankruptcy frequency in the estimation sample. An increase in the risk at large enterprises leads to a (marginal) decrease in the risk at (the large number of) small enterprises, so that the average probability is unchanged.

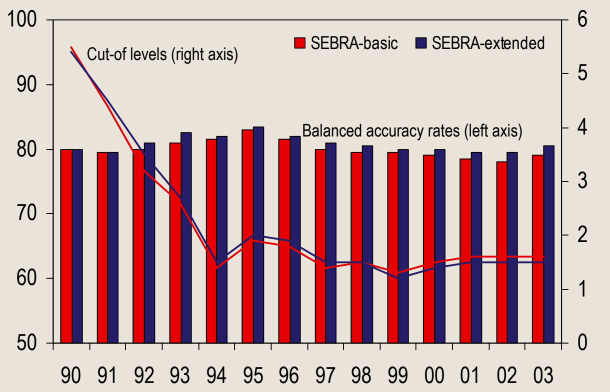
SEBRA Extended

The extended version is the same as the basic version but also includes variables for trade accounts payable, government taxes payable and size. These variables are either directly or indirectly related to an enterprise's size. The dividend variable is excluded from both of the new variants of the model.

Data and methodology[8]

We use key financial figures based on enterprises' annual accounts and information on their age, size and industry classification to estimate the models. In principle, all Norwegian non-financial limited companies with total assets in excess of NOK 500 000 are included in the sample. However, some enterprises drop out as a result of accounting shortcomings. The estimation period is from 1990 to 2002. The variable that is explained is defined by the coincidence of the events: "Enterprise stops filing accounts the following year" and "Bankruptcy filed". In around 20 per cent of cases, bankruptcy is filed three years after the last set of accounts is submitted. This means that the model can only be tested and re-estimated on accounts two to three years ahead of the last available set of accounts. In all, there are about a million sets of annual accounts in the estimation sample, of which around 20 000 represent bankruptcy observations. As in the original SEBRA model, we use a generalised logit model to estimate the probability of an enterprise filing for bankruptcy.[9]

Chart 2 Balanced accuracy rates and cut-off points. Per cent. Annual figures. 1990-2003



5. Accuracy of the SEBRA models

Accuracy at enterprise level

When evaluating bankruptcy prediction models, it is normal to determine a cut-off level for predicted bankruptcy probabilities, so that all observations above this level are classified as bankrupt, and all those below this level are classified as non-bankrupt. The cut-off level can, for example, be set in a way that the proportions of correctly predicted bankruptcy and non-bankruptcy observations are the same for both variables (balanced accuracy) (see Chart 2).

Accuracy rates are consistently lower for SEBRA Basic than for SEBRA Extended, but the differences are small.[10] This does not mean that the different variants of the model assign each enterprise the same bankruptcy probability or identify the same bankruptcies. The estimates for individual enterprises can be very different. Accuracy rates for SEBRA Extended are approximately the same as for the original model.

Which of the two new versions of the model best approximates actual loss probabilities depends on how good a substitution variable bankruptcy is for defaults and losses. If bankruptcy is viewed as a good substitution variable for both small and large enterprises, we should attach the most weight to the classification in SEBRA Extended. Otherwise, we should attach the most weight to the classification in SEBRA Basic.

Charts 3 and 4 show average bankruptcy probabilities and actual bankruptcy frequencies for 1990 and 2002 for enterprises divided into eight risk groups on the basis of high or low bankruptcy probability (see Table 2 for the distribution criteria). We have chosen 1990 and 2002 because these are the first and last years in the estimation sample, but equivalent results are obtained for all of the years in the sample. There is generally a good match between predicted bankruptcy probabilities and actual bankruptcy frequencies for the different risk groups throughout the estimation period.

Table 2. Proportion of bank debt in different risk groups. Per cent.

Risk group	1	2	3	4	5	6	7	8
Bankruptcy probability (P), percentage	*P > 20*	*20 > P > 10*	*10 > P > 5*	*5 > P > 2*	*2 > P > 1*	*1 > P > 0.5*	*0.5 > P > 0.1*	*P < 0.1*
SEBRA Basic	0.01	0.10	0.35	3.60	4.74	15.96	66.81	8.44
SEBRA Extended	0.01	0.06	0.21	0.93	1.83	4.42	23.75	68.79

[8] A technical paper presenting the new variants of the model in greater detail will be available at a later date.

[9] The method is classified as a parametric generalised additive model (GAM). This model is described in Bernhardsen (2001) and Eklund et al. (2001). Berg (2007) estimates a non-parametric GAM for bankruptcies in Norway based in part on key figures from SEBRA.

[10] The models' accuracy can also be evaluated for all cut-off levels using an ROC analysis. The accuracy rates for bankruptcy and non-bankruptcy observations are plotted against one another, and the area under the resulting curve is calculated. A completely arbitrary classification will give an ROC value of 50 per cent for large samples, while a value of 100 per cent shows perfect classification. The ROC values for SEBRA Basic and SEBRA Extended are 88 and 89 per cent respectively.

Chart 3 Pedicted probability of bankruptcy and actual bankruptcies in various risk groups. 1990

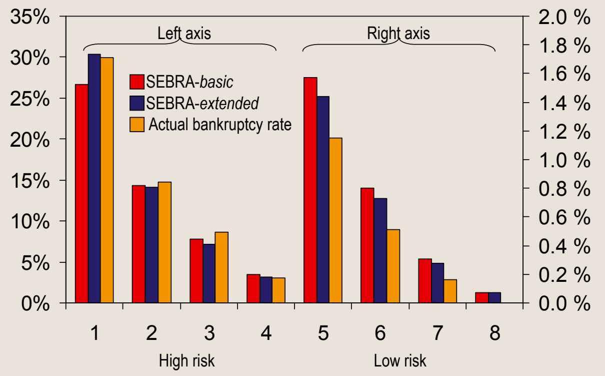
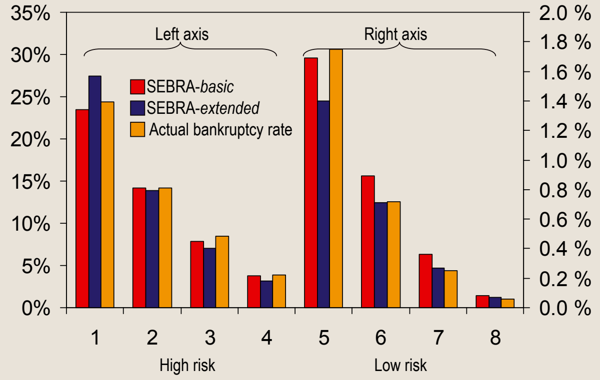


Chart 4 Pedicted probability of bankruptcy and actual bankruptcies in various risk groups. 2002



Accuracy at aggregate level

The analysis above indicates that the two new SEBRA models have good predictive power at enterprise level. The charts also suggest that the differences between the two model variants are small. The differences between the models are larger when we weight bankruptcy probabilities with the amount of debt at each enterprise (see Table 2). We see here that most bank debt is in the low-risk groups in both models. However, as large companies are assigned a higher bankruptcy probability in SEBRA Basic, a larger share of bank debt is in the high-risk groups in this model.

Chart 5 shows the average predicted bankruptcy probabilities for the two models and actual bankruptcy rates for each year in the estimation period. Actual bankruptcies are represented by the last set of accounts submitted for enterprises that go bankrupt, hereafter referred to as bankruptcy accounts. Up to three years can elapse between the last set of accounts being submitted and bankruptcy being filed. This means that, in the last available accounts year t, we can only perform a complete count until year t-3. (Thus, for example, with 2006 data available, we can count which accounts in 2003 are bankruptcy accounts.) With the exception of 1992 and 2000–2001, when the predicted bankruptcy probabilities are higher and lower respectively than actual bankruptcy accounts, there is close accord between predicted and actual bankruptcy accounts. This indicates that both SEBRA models are successful in predicting aggregate bankruptcy rates in the enterprise sector.

Banks' recorded loan losses are determined by the size of bad loans (potential loan losses) and the proportion of each bad loan actually lost (loss given default). We do not have information on bad loans at enterprise level and so cannot measure potential loan losses directly. However, we know that they will be larger than the amount of debt in bankruptcy accounts, because banks will also have losses on loans to enterprises that do not go bankrupt. We can also add up the debt in all terminal accounts – in other words, the accounts of all enterprises that go bankrupt, are wound up for some other reason, or are taken over (see Chart 6). Many of the enterprises that stop filing statements without going bankrupt settle their debt before being wound up or taken over. On the other hand, there may also be losses on loans to enterprises that continue operations. In our

Chart 5 Acutal banktruptcy accounts and average predictions. Per cent of total and probabilities. 1990-2003

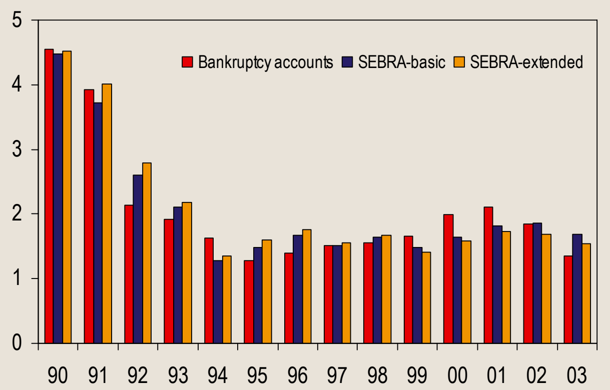


Chart 6 Debt in terminated and bankrupt firms. Billions of NOK. 1999-2002

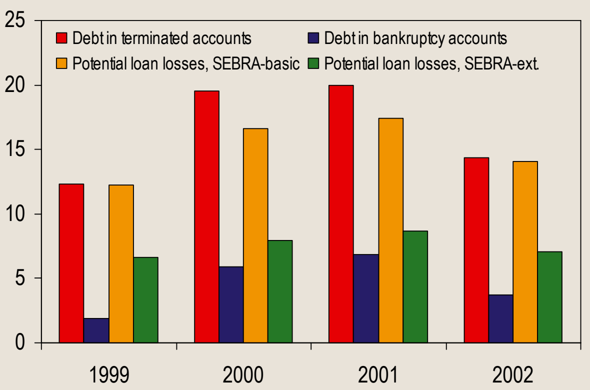
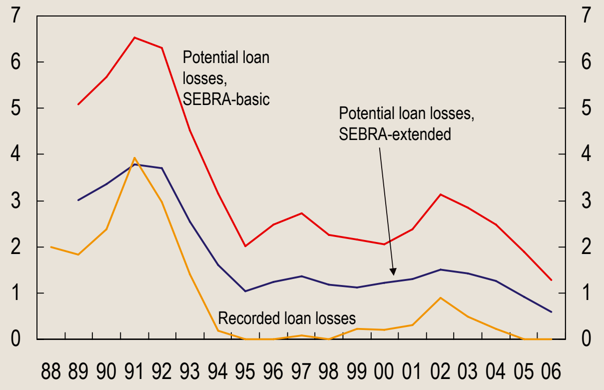


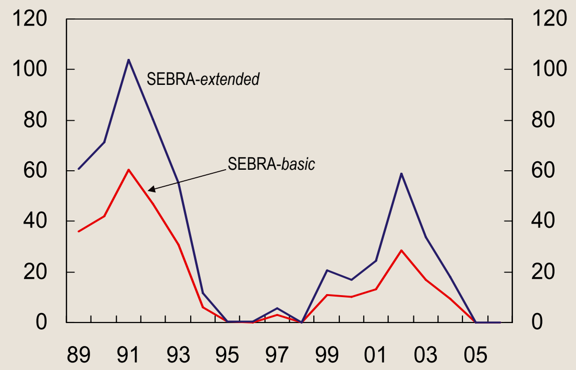
Chart 7 Estimates of potential loan losses and banks' recorded loan losses.1) Per cent of lending. Annual figures. 1988-2006



1) Potential loan losses are defined as expected losses with a 100% loss given default

Source: Norges Bank and Statistics Norway

Chart 8 Implied measures for loss given default. Per cent of defaulting loans. 1989-2006



Source: Norges Bank

opinion, potential loan losses are closer to the sum of the debt in all terminal accounts than to debt in bankruptcy accounts alone.

By weighting the bankruptcy probabilities with debt at each enterprise and then adding all of the enterprises together, we obtain an estimate of expected potential loan losses due to bankruptcy. To take account of actual losses being higher than losses due to bankruptcy, we have chosen to scale up the bankruptcy probabilities.[11] In Chart 6, we show expected potential loan losses following this upscaling. We see that the estimates from SEBRA Basic are close to total debt in bankruptcy and terminal accounts, while the estimates from SEBRA Extended are relatively close to the debt in bankruptcy accounts. Since debt in bankruptcy accounts represents an absolute minimum for potential loan losses, we have greater confidence in the estimates from SEBRA Basic.

Chart 7 presents estimates of potential loan losses from the two models and banks' recorded loan losses.[12] In banks' accounts, recorded loan losses are calculated as changes in loss provisions plus new losses less recoveries on loans previously written off.[13] For our purposes, it is more appropriate to consider recorded loan losses as a product of the size of bad loans and the proportion of each bad loan that is not recovered (loss given default). By dividing recorded loan losses by estimates of potential loan losses, we obtain a measure of loss given default at macro level (see Chart 8). In the literature, loss given default in different countries is rarely reported to be higher than 60 per cent and rarely lower than 10 per cent.[14] Based on this, loss given default from SEBRA Basic seems more realistic than that from SEBRA Extended. The reason for loss given default in Chart 8 being estimated at zero in some years is recoveries on loans previously written off. A better match with movements in recorded loan losses over time is also obtained with SEBRA Basic (see Chart 7).

The main reason why SEBRA Basic is more accurate in terms of both the level of and changes in banks' aggregate loan losses is that it attaches less weight to the enterprise's size (see discussion above). This suggests that we should use SEBRA Basic when projecting banks' loan losses. However, since it is slightly less accurate when it comes to bankruptcies at enterprise level, we should use SEBRA Extended instead for analyses where the emphasis is more on bankruptcies than on aggregate loan losses.

6. Projecting banks' loan losses

The key figures in the original SEBRA model can be projected using macroeconomic scenarios from Norges Bank's macro models (see Frøyland and Larsen (2002)). This makes it possible to calculate estimates of potential loan losses in the future. Such estimates can be produced both for a baseline scenario and for various stress test scenarios.

Norges Bank is currently further developing the models for projecting and stress testing banks' losses on loans to enterprises. Use of SEBRA Basic will make it easier to project enterprises' accounts, because we need only project the items included in the calculation of the basic key figures for earnings, liquidity and financial strength. The changes in the model and projection tool will probably result in better estimates of banks' loan losses.

In this work on further developing projections and stress tests, we have found that loss given default can be projected accurately using a simple dynamic model where changes in commercial property prices are included as an explanatory factor. This is not surprising

[11] There is a limited basis for how best to scale up the probabilities of bankruptcy from SEBRA into probabilities of loss or default. A factor of 2 was estimated in a simple statistical model for mis-classification as in Bernhardsen (2001).

[12] We have lagged the model estimates by one year here. This is intuitive because the bankruptcy probabilities are calculated on the basis of enterprises that have not yet gone bankrupt (see the definition of the bankruptcy event in Section 4).

[13] See pages 31–32 of Financial Stability 2/01 for a more detailed discussion of banks' loan losses and loss provisioning practice.

[14] See, for example, Dermine and de Carvalho (2006).

as banks' lending to enterprises is often secured against commercial property. A reduction in the value of the collateral gives banks poorer cover for the balance on a bad loan when the collateral is sold. It is also likely that other types of security may be closely correlated with commercial property prices.

In *Financial Stability* 2/06, we estimated a relationship for loss given default based on the original SEBRA model.[15] According to the estimated model, a 10 per cent drop in commercial property prices leads, in isolation, to an increase in loss given default of around 11 percentage points. Loss given default also tends towards a constant level of 35 per cent over time. Given actual movements in commercial property prices, dynamic estimates for loss given default show very good approximations both two and three years ahead. This indicates that we can produce good estimates of banks' loan losses provided that we are able to project the key figures in the model.

7. Summary

We have discussed various reasons for further developing the SEBRA model. The most important reasons are to improve estimates of banks' loan losses and to obtain a model which makes it easier to make projections and perform stress tests. We have estimated and tested two new versions of the SEBRA model: SEBRA Basic and SEBRA Extended. These two versions of the model are, respectively, a simplification and a refinement of the original model. SEBRA Basic has a marginally lower accuracy rate than SEBRA Extended for bankruptcies at enterprise level, but is better suited to estimating banks' potential loan losses. Furthermore, the basic version is easier to project using different scenarios for macroeconomic developments. We have shown that SEBRA Basic provides good estimates of banks' recorded loan losses. In the future, we will use SEBRA Basic in analyses of banks' loan losses, but SEBRA Extended in analyses where the emphasis is more on bankruptcies than on aggregate loan losses. Norges Bank will continue its work on further developing the projection and stress test module for banks' losses on loans to enterprises.

References:

Berg, D. (2007): "Bankruptcy prediction by generalized additive models". *Applied Stochastic Models in Business and Industry*, Vol. 23, No. 2, pp. 129–143

Bernhardsen, E. (2001): "A model of bankruptcy prediction". *Working Paper* 2001/1, Norges Bank

Bernhardsen, E. and K. Larsen (2002): "Banks' pricing of risk associated with corporate lending". *Economic Bulletin* 1/2003, Norges Bank

Dermine, J. and N. de Carvalho (2006): "Bank loan losses-given-default". *Journal of Banking and Finance*, Vol. 30, Issue 4, pp. 1219-1243

Hagen, J., A. Lund, K.B. Nordal and E. Steffensen (2005): "IMF stress testing of the Norwegian financial sector". *Economic Bulletin* 4/2005, Norges Bank

Eklund, T., K. Larsen and E. Bernhardsen (2001): "Model for analysing credit risk in the enterprise sector". *Economic Bulletin* 3/2001, Norges Bank

Frøyland, E. and K. Larsen (2002): "How vulnerable are financial institutions to macroeconomic changes? An analysis based on stress testing". *Economic Bulletin* 3/2002, Norges Bank

Larsen, K. and K.M. Bjerkeland (2005): "Are unexpected loan losses lower for small enterprises than for large enterprises?" *Economic Bulletin* 3/2005, Norges Bank

Norges Bank, *Financial Stability* 2/06

[15] Loss given default (t) = 0.085 + 0.76loss given default (t-1) – 1.09 Δln(commercial property prices).