Bayesian Methods for Improving Credit Scoring Models

Gunter Löffler, Peter N. Posch*, Christiane Schöne

First Version: January 2004. This Version: 31st May 2005

Department of Finance, University of Ulm, Helmholtzstrasse 18, 89069 Ulm, Germany.

* Corresponding author. Tel.: ++49-731-5023596; Fax.: ++49-731-5023590 ;
E-Mail address: pposch@uni-ulm.de (P.N. Posch)

Abstract

We propose a Bayesian methodology that enables banks to improve their credit scoring models by imposing prior information. As prior information, we use coefficients from credit scoring models estimated on other data sets. Through simulations, we explore the default prediction power of three Bayesian estimators in three different scenarios and find that they perform better than standard maximum likelihood estimates. We recommend that banks consider Bayesian estimation for internal and regulatory default prediction models.

Keywords: Credit Scoring, Bayesian Inference, Bankruptcy Prediction
JEL-Classification: C11, G21, G33

1 Introduction

Banks use credit scoring models when approving and pricing loans. Under the proposed new Basel capital accord (Basel II)[1], banks can also use their default probability estimates for calculating regulatory capital. Therefore, accurate default prediction is in the interest of banks as well as regulators.

It is generally easier to predict defaults accurately if a large data set (including defaults) is available for estimating the prediction model. This puts not only small banks, which tend to have smaller data sets, at disadvantage. It can also pose a problem for large banks that began to collect their own historical data only recently, or banks that recently introduced a new rating system.

We propose a Bayesian methodology that enables banks with small data sets to improve their default probability estimates by imposing prior information on the estimates. As prior information, we use coefficients from credit scoring models estimated on other data sets. In many cases, such prior information will be readily available from the academic literature (e.g. Altman (1968) or Shumway (2001)). It could also be made available by regulators. The Deutsche Bundesbank or the Banque de France, for example, have large data sets with corporate financial statements which they use to estimate credit scoring models (see Engelmann et al. (2003) and Banque de France (2001)).

For illustrating the accuracy gains from Bayesian estimation, we use simulations based on a data set comprising the non-financial firms in the S&P 1500 index. Within a logit estimation framework we estimate the coefficients of a credit scoring function with standard maximum likelihood ("straight logit" hereafter) and compare them to an approximate Bayes, an empirical Bayes and a Stein rule estimator. In order to evaluate the quality of the resulting default probability estimates, we use the accuracy ratio and the Brier score, two measures commonly applied in the literature.

In our settings, all three Bayesian estimators are significantly more accurate than the straight logit estimator. We therefore recommend that banks use a Bayesian estimator for their internal and regulatory default prediction models.

A closely related paper is Frerichs and Wahrenburg (2003). The simulations of the

[1]See Basel Committee on Banking Supervision (2003).

authors suggest that pooling of estimation samples might improve credit rating systems. As in our paper, the benchmark estimation model is a logistic credit scoring function. The key difference between Bayesian methods and pooling is that the latter is only possible if full access to additional data sets is available. Such data, however, will usually not be available for free, and in some cases they may not be available at all. Another advantage of the Bayesian method is that it provides a natural way for dealing with structural differences between a bank's internal data and additional, external data. In practice, the true scoring function may differ across the data sets, the small internal data set may contain information that is missing in the larger external data set, or the variables in the two data sets are not exactly the same but related. In each of these cases pooling is feasible, but it requires the modeler's explicit decision on how to deal with the structural differences, e.g. restricting coefficients to be equal across the data sets or determining a way of imputing missing values. Bayesian methods by contrast endogenously combine in-sample and prior information according to their precision; structural differences are accounted for as they affect the relative precision of prior and in-sample information.

Seminal contributions to default prediction are Altman (1968) and Beaver (1968), recent ones include Shumway (2001) and Chava and Jarrow (2004). Statistical methods for evaluating default probability estimates are discussed in Sobehart and Keenan (2001), Engelmann et al. (2003), Stein and Jordao (2003) and Stein (2005). Stein and Jordao (2003) and Stein (2005) measure the power of scoring models using accuracy ratios, and attach a monetary (i.e. dollar) value to a bank's application of a model that is more powerful.

The remainder of this paper is organised as follows. Section 2 introduces the data set. Section 3 presents the Bayesian estimators. Section 4 describes the accuracy measures that we use for evaluation. Section 5 describes the simulation set-up. Section 6 presents the results, and Section 7 concludes.

2 Data

Following Altman (1968), we use five explanatory variables for default prediction: working capital/total assets (WC/TA), retained earnings/total assets (RE/TA), earnings before interest and taxes /total assets (EBIT/TA), market value of equity/book value of total liabilities (MV/TL), and sales/total assets (S/TA). For an economic interpretation of these variables, see Altman (1968), pp. 594 - 596. In order to simulate a scenario in which the external and internal data sets contain different variables, our data set additionally contains book value of equity/book value of total liabilities (BV/TL).

We obtain data on these variables for the S&P 1500 index companies from Worldscope. We collect the data for the end of each year from 2000 to 2004 and exclude financial firms and observations with missing values. Furthermore, we winsorise (see e.g. Barnett and Lewis (1994)) each variable symmetrically to a 99% confidence band, as is done in Shumway (2001). Table 1 provides summary statistics for the remaining 4,558 observations.

===Insert table 1 around here.===

Expected default probabilities are calculated using the coefficients estimated in Shumway (2001), Table 2. In our data set, this leads to a mean expected default probability of 0.44% with a standard deviation of 1.15%, a minimum of 0% and a maximum of 19.73%. This compares to a default rate of 0.81% in the data used by Shumway (2001).

3 Bayesian Estimators

The Bayesian estimators we propose are based on Adkins and Hill (1996), who show how prior information in the form of a coefficient vector can be used in a Probit analysis. We apply their proposal to logit estimation, which is very similar to Probit, but more common in the default prediction literature.

The estimators discussed by Adkins and Hill (1996) can be classified into two types. The first type (Type-I) weighs the prior information vector equally for all vector entries, i.e. each prior coefficient has the same, proportional influence on the posterior (Bayesian) coefficient estimates. In the second type (Type-II), the weights use information on the variance of the prior coefficient vector without assuming a particular structure of its variance.

The general form of the Type-I Bayesian Estimators is given by equation (1), where β_p denotes the prior information vector, which is obtained independently of the given sample, while $\widetilde{\beta}$ denotes the maximum likelihood (ML) estimation vector of the (unknown) coefficients β in the current analysis (i.e. 'Straight logit estimator') and $w \in [0,1]$ is a given weight.

$$\beta_{Bayes} = w \cdot \beta_p + (1-w) \cdot \widetilde{\beta} \tag{1}$$

As in Adkins and Hill (1996) we employ an empirical Bayes estimator (EBE) and the (James-)Stein rule estimator (SRE), proposed by James and Stein (1961). For details on the close connection between these estimators see Judge et al. (1985), pp. 117-121.

Assuming that the covariance matrix of the prior is equal to the covariance matrix of $\widetilde{\beta}$ times a constant factor c results in the weight $w = (1+c)^{-1}$. Using the marginal distribution of the estimator in the current sample conditional upon the prior information (β_p) leads to the weights of the empirical Bayes estimator (see Adkins and Hill (1996) for a detailed derivation):

$$w_{EBE} := \frac{J-2}{(\widetilde{\beta}-\beta_p)' I(\beta) (\widetilde{\beta}-\beta_p)} \ , \tag{2}$$

where J denotes the number of restrictions imposed through the prior, and $I(\beta)$ is the information matrix, i.e. the inverse of the covariance matrix of β. It is estimated using the covariance matrix of $\widetilde{\beta}$. If the point estimates of the coefficient vector β have a small variance, the weight w_{EBE} is relatively small, and hence the information in the current sample is weighed more heavily.

The weight for the SRE is given by

$$w_{SRE} := \frac{J-2}{2\left(\ln L(\beta_p) - \ln L(\widetilde{\beta})\right)} \ , \tag{3}$$

where $\ln L(\cdot)$ is the Log-Likelihood of $(\cdot)$. Here the prior information is weighed heavily when sample and non-sample information agree, i.e. if the Log-Likelihood functions of the prior and the current sample coefficients do not differ much.

To avoid excessive shrinkage ('overshrinkage'), we set the weight to one if $w > 1$. In these cases, the prior is used as the posterior (rather than the Bayesian estimator).

By contrast, Type-II Bayesian Estimators require an estimate of the covariance matrix of the prior coefficient vector. The approximate Bayes estimator is given by

$$\beta_{ABE} = [A + I(\beta)]^{-1} \left(A\beta_p + I(\beta)\widetilde{\beta} \right) \tag{4}$$

where A is the information matrix (i.e. the inverse of the covariance matrix) of the prior distribution. Following Zellner and Rossi (1984) we estimate A within the current sample by imposing the prior coefficient vector β_p as restriction. We obtain A as the negative Hessian matrix of the logistic function at the prior coefficient vector β_p (see e.g. Greene (2003) for a derivation of the Hessian). The information matrix $I(\beta)$ is again estimated using $I(\widetilde{\beta})$.

In contrast to the EBE and the SRE, the ABE allows for differences in the precision of prior coefficients. This difference is particularly important when the internal data set lacks variables that are contained in the external data set or vice versa. In such situations there are at least two different approaches to estimating Bayesian coefficients.

The bank could replace the missing variable with another variable that is correlated with the missing one. For example, the external data set may contain data on market value of equity, whereas the bank has only data on book value of equity. In such a scenario, one could use the estimated coefficient for market value as the prior when estimating a Bayesian coefficient for book value. In section 6.2 we demonstrate Bayesian estimation in this case using the prior coefficients of market value to derive estimators for book value.

The second approach is a restricted estimation within the bank's own data set. If, for example, variable 3 is missing in the external data set, we would estimate its coefficient in the internal data set, restricting the remaining coefficients to the prior coefficients.[2] The resulting restricted coefficient vector $\beta_r = (\beta_r^0, \beta_p^1, \beta_p^2, \beta_r^3, \beta_p^4, \beta_p^5)'$, where β_x^i denotes the i-th entry of the vector β_x, is then used as the prior. In section 6.2 we use the two most significant variables EBIT/TA and MV/TL to demonstrate this estimation procedure.

4 Accuracy Measures

We compare the accuracy of the default predictions under straight logit, approximate Bayes, empirical Bayes, and Stein rule estimations using the accuracy ratio (see Sobehart

[2] The constant is left unrestricted in the estimation.

and Keenan (2001) or Engelmann et al. (2003)) and the Brier score (Brier (1950), Frerichs and Wahrenburg (2003), Grunert et al. (2005)).

The calculation of the accuracy ratio (AR) is based on the cumulative accuracy profile (CAP). To construct the CAP all debtors are sorted according to their score, starting with the debtor with the lowest score (highest probability of default) down to the debtor with the highest score. A CAP is then obtained by plotting the proportion of defaulted debtors against the proportion of all debtors, see figure 3.

===**Insert figure 3 around here.**===

A 'perfect' scoring would assign the lowest score to all defaulters and higher scores to all non-defaulters. The corresponding CAP increases linearly until all defaulters are included and then stays at 100%. In constrast, a non-informative scoring would randomly assign scores. In such a random scoring we expect x% defaulters among the x% of all debtors with the lowest score and accordingly for higher scores. The CAP of this rating is linear with a slope of one, starting at the origin. Real scoring models are between these extremes. The accuracy ratio is defined as the area between the CAP of the analyzed scoring system and the non-informative system (area A in figure 3) divided by the area between the CAP of the 'perfect' scoring model and the CAP of the non-informative scoring model (area B in figure 3). A scoring model with high discriminative power has an accuracy ratio close to 100%, while the minimum value of the AR is 0% for the random scoring model.

The Brier score combines the quality of the ranking with the accuracy of the estimated probabilities of default. It is defined as follows:

$$Brier = \frac{1}{N} \sum_{i=1}^{N} (PD_i - I(Default_i))^2 \tag{5}$$

where PD_i is the estimated default probability, and $I(Default_i)$ is an indicator variable that takes the value 1 if firm i defaults and zero otherwise. Thus, the Brier score is the mean squared error of PD_i.[3]

[3] In the setting of this paper we could calculate the mean squared error using the correct default probability instead of the indicator variable $I(Default_i)$ because it is set at the start of the simulation.

5 Simulation Design

With the simulation, we demonstrate how a bank with a small debtor data set can improve the accuracy of its default probability estimates. Such a bank could, for example, obtain coefficient estimates from an academic study, a larger bank, a rating agency, or from regulatory bodies. These estimates can be used as prior information in a Bayesian estimation of the posterior coefficient vector.

===Insert figure 1 around here.===

We simulate both the external information and the bank's hypothetical data sets. The simulation, which is summarised in Figure 1, is structured as follows: We obtain a large 'external' data set by drawing random samples with replacement (bootstrapping) from the initial data set described in section 2. Considering the data set used by Shumway (2001) to be representative with respect to the number of observations we expand our data set such that it is 6.4 times larger than the initial one, yielding 29,500 firm-years.[4] The expansion is done anew in each repetition n.

Using the coefficient vector as given in Shumway (2001), we then calculate expected default probabilities for each observation in the expanded data set. We simulate defaults based on the expected default probability using a uniformly distributed random variable. Prior coefficient estimates β_p are obtained using the expanded data set with maximum likelihood. Therefore, the prior information is different in each repetition n. Note that this coefficient vector is the only prior information needed to calculate Bayesian estimates. In practice, it could come from an external source such as another bank or an agency.

Afterwards we draw without replacement from the expanded data set (including the defaults) to obtain a smaller 'internal' data set of size S. This corresponds to the internal data set of a bank. Using this small data set, we first run a straight logit estimation

Since the correct default probability is not available in practical applications, this 'modified' Brier score is not used here.

[4] The resulting data set is quite small compared to the Deutsche Bundesbank database which contains over 325,000 balance sheets (Engelmann et al. (2003)) and the FIBEN database of the Banque de France with 180,000 balance sheets (Banque de France (2001)).

(SLE). This is what a bank can achieve without any outside information. In the next step we calculate the ABE, SRE and EBE.

Then, $K = 100$ new default vectors are generated randomly within the small sample of size S. This results in 100 new 'samples' and represents an out-of-time setting, in which the performance of the straight logit, approximate Bayes, empirical Bayes, and Stein rule estimators is tested. Accuracy measures are saved and used to evaluate the accuracy of the Bayesian estimators relative to the straight logit estimator. Increasing the number of out-of-sample repetitions K when simulating the defaults has no noteworthy impact on the calculated measures of accuracy. For the further analysis, we use the means of the accuracy measures across the $K = 100$ results. We assess the simulation error through a Wilcoxon matched-pairs test for equality of distributions. The reported differences between Bayesian estimates and straight logit are all significant on a level of less than 0.01% if 1,000 simulation steps are conducted. We therefore repeat the calculations $N = 1,000$ times.

Figure 2 illustrates the effects of the Bayesian estimation for EBIT/TA. The figure shows the empirical density distribution of the 1,000 simulated coefficients. The prior coefficient varies in each repetition because we randomly draw the external data set from which it is computed. Its variation is smaller than the variation of the SL because the internal data set on which the latter is based is smaller. The variation of the Bayesian estimators is in between as they combine the prior with the straight logit.

===**Insert figure 2 around here.**===

The size S of the smaller data set is fixed at the outset of the simulation. We vary its proportion of the large data set between $S = 5\%$ and $S = 10\%$, corresponding to 1,475 and 2,950 firm-years, respectively.[5]

[5] For $S = 1\%$ there are typically too few defaults in the internal data set (the expected number of defaults is 1.3). If the number of defaults in a given simulation step is insufficient to estimate the logit model we repeat that simulation step until we get sufficient defaults.

6 Results

In the remainder of this section, the performance of the three Bayesian estimators is compared to that of the straight logit estimator. As a benchmark table 2 reports the accuracy ratio when using the prior coefficient vector. Recall that the prior is the maximum likelihood estimator for the *expanded* data set. In our setting this provides efficient estimates for the *smaller* data set because the latter is a subset of the expanded data set. The accuracy ratio achieved by these estimates should be close to the upper boundary for the accuracy obtainable by any estimation procedure.[6]

===**Insert table 2 around here.**===

We first report the results for the case in which there are no structural differences between the two data sets. Then we examine situations in which the variables in the internal and external data set differ. The results for the Brier score are qualitatively the same as for the accuracy ratio and therefore not discussed. Key results are reported in the Appendix, details are available upon request.

6.1 No structural differences between internal and external data

Table 3 records the simulated accuracy ratio of the three Bayesian estimators and the straight logit approach and compares the former to the latter one.[7]

The accuracy ratio of the three Bayesian estimators is about 5 percentage points (pp) higher in the $S = 5\%$ setting and about 2.5 pp in the $S = 10\%$ setting. The approximate Bayes estimator (ABE) performs best with a mean increase in the accuracy ratio by 5.7 pp for the $S = 5\%$ setting and 2.7 pp in the $S = 10\%$ setting, compared to a straight logit estimation (SLE). Comparing the performance of the two best-performing Bayesian estimators, the approximate Bayes versus the empirical Bayes estimators, the approximate Bayes performs significantly better according to the Wilcoxon test.

[6] An efficient coefficient vector does not necessarily maximize the accuracy ratio.

[7] Overshrinkage occurs less often for the Stein rule estimator (24% in the simulation with $S = 10\%$ and 29.5% in the $S = 5\%$-simulation) than for the empirical Bayes estimator (33.4% in the $S = 10\%$-simulation and 33.8% in the $S = 5\%$-simulation).

The 'Count' is defined as the fraction of simulation repetitions in which the mean accuracy ratio of the Bayesian estimator is higher than the mean accuracy ratio of the straight logit estimator. It can be interpreted as follows: assume there are two sets of the same size of mutually independent banks, and one set implements Bayesian estimation for its credit scoring models, while the other set continues to use a standard logit approach. The average of the accuracy ratios in the Bayesian bank set will be higher than that of the other set in at least 'Count' of the cases.

===**Insert table 3 around here.**===

6.2 Structural Differences

In the previous section we examined the ideal case where the structure of the prior data set and the random subsample are identical. Now we impose structural differences. We first assume that data for EBIT/TA or MV/TL are missing in the external data set.

We estimate the missing variable's coefficient within the bank's own data set using a restricted ML estimation as described in section 3. If EBIT/TA is missing, for example, we restrict the coefficients on WC/TA, RE/TA, MV/TL and S/TL to be equal to the prior and leave the coefficient on EBIT/TA unrestricted.

Table 4 records the descriptive statistics for the accuracy ratio and compares the three Bayesian estimators to the straight logit in the case of $S = 5\%$ and $S = 10\%$. The Bayesian estimators improve the default probability estimates by 2 to 4 pp as measured by the accuracy ratio. In this case the approximate Bayes and the empirical Bayes perform equally well on average in the $S = 10\%$ setting, while the approximate Bayes is significantly better in the $S = 5\%$ setting.

===**Insert table 4 around here.**===

===**Insert table 5 around here.**===

When MV/TL is missing in the prior data set, the Bayesian estimators improve the default probability estimates by 1.9 to 4.8 pp as measured by the accuracy ratio (see table 5). Again, the approximate Bayes estimator performs best with 4.8 pp in the $S = 5\%$ and 2.4 pp in the $S = 10\%$ setting, followed by the empirical Bayes estimator (4.3 pp improvement in the $S = 5\%$ and 2.3 pp in the $S = 10\%$ setting).

The second approach described in section 3 concerning structural differences is the replacement of the missing variable with another variable which is correlated with the former. We demonstrate this approach by replacing market value (MV) by book value (BV) in the small data set. The correlation between MV/TL and BV/TL is 0.823 showing that book value is a good proxy for market value.[8]

The results for the $S = 5\%$ and $S = 10\%$ analysis are reported in table 6. In this setting the relative performance of the Bayesian estimators is much better than in other settings. The accuracy ratio increases by 6.5 to 9.4 pp ($S = 5\%$) and 3.3 to 4.5 pp ($S = 10\%$).

The advantage is larger than in the base case of section 6.1 because the prior is based on the same information as in the base case whereas the straight logit uses only a proxy for market value and thus less information. The advantage is even larger when compared to the cases examined above where EBIT/TA or MV/TL are missing in the expanded data set. In those cases the prior lacks information that is incorporated in the straight logit.

===**Insert table 6 around here.**===

6.3 Interpretation of Results

The results in Stein and Jordao (2003) and Stein (2005) suggest that a 2 to 4 pp difference in the accuracy ratios is economically significant, i.e. a bank would benefit from applying the scoring system with the higher accuracy ratio. In our analysis Bayesian estimators achieve improvements from 2 to 9 pp. This improvement is obtained even with

[8]The multiple correlation coefficient between MV/TL on the one hand and WC/TA, RE/TA, S/TL on the other hand is 0.524, so book value adds valuable information not contained in the other variables.

complication such as missing variables.

In the base case the mean accuracy ratio of Bayesian estimators comes close to the mean accuracy ratio achievable in the ideal case in which coefficients can be estimated on the full data set and Bayesian estimation is obsolete (see Panel A of table 3 and table 2). This suggests that Bayesian estimators effectively combine the available information.

Comparing the Bayesian estimators to each other the Stein rule estimator performs less well than the approximate Bayes and the empirical Bayes estimators. Among the latter two, the ABE performs better than the EBE in most settings and equally well in the remaining settings. The ABE accounts for differences in the precision of coefficients, which are generally present in our setting as the variables differ in their significance, and which are increased when imposing structural differences. Since differences in precision are likely to be present in practical applications we recommend the use of the approximate Bayes estimator.

Due to lack of data, we cannot directly explore the benefits of Bayesian estimation in a situation that could often arise in practical applications of our methodology: while many banks use qualitative assessments of management quality, business risk or other factors in their credit scoring process, qualitative information is usually not available in an external data set, and so there will typically be no prior for it. This situation, however, corresponds to the case examined in section 6.2, where we assumed that a variable available in the small data set is missing in the large data set. In our simulations, leaving MV/TL out of the logit regression reduces McFadden's R^2 on average from 22.2% to 9.5%. Grunert et al. (2005) use Probit models to examine the default prediction power of internal rating data of major German banks and find that the R^2 decreases from 36.0% to 26.8% when the qualitative component of the bank's rating is left out. It thus appears that the problem of missing MV/TL studied in this paper is more severe than the problem of missing qualitative factors, suggesting that Bayesian estimators might perform well in the latter case, too.

7 Conclusion

The purpose of this paper is to determine whether a bank can improve the accuracy of its default probability estimates using Bayesian inference.

The accuracy of Bayesian estimators is evaluated in comparison to straight logit estimators. As the Bayesian estimators incorporate prior information according to its precision, one expects them to perform better than a straight logit estimator. The results of our analysis support this conjecture. On average, accuracy ratios of Bayesian estimates are 2 to 9 percentage points higher than accuracy ratios of standard logit estimates. The improvement is achieved even when an important predictive variable is missing in the prior data set or if one has to resort to proxy variables.

As a result, we recommend that financial institutions implement the estimation process proposed in this paper. The application of this method is not confined to the case where the prior information comes from an external source. A bank expanding in a new market segment, for example, could use its own data from established segments as a prior for the scoring model to be used for the new segment. We emphasise that the prior information is just a vector of coefficients from a credit scoring model, i.e. information that is often readily available. Note, too, that the implementation of the Bayesian estimators does not require any judgmental decisions apart from choosing the source of the prior information.

1. Draw large random sample with replacement (source of prior information)
2. Generate random defaults
3. Obtain prior information through logit estimation (using data generated in steps 1 and 2)
4. Draw random subsample from the initial data set (without replacement) of size S
5. Calculate straight logit (using data from step 4) as well as Bayesian estimates (using data from step 4 and priors from step 3)
6. Generate random defaults to evaluate default probability estimates out-of-time. Repeat this evaluation $k = 1, ..., K = 100$ times
7. Repeat steps 1.-6. $n = 1, ..., N = 1,000$ times

Figure 1: Overview of the Simulation Process

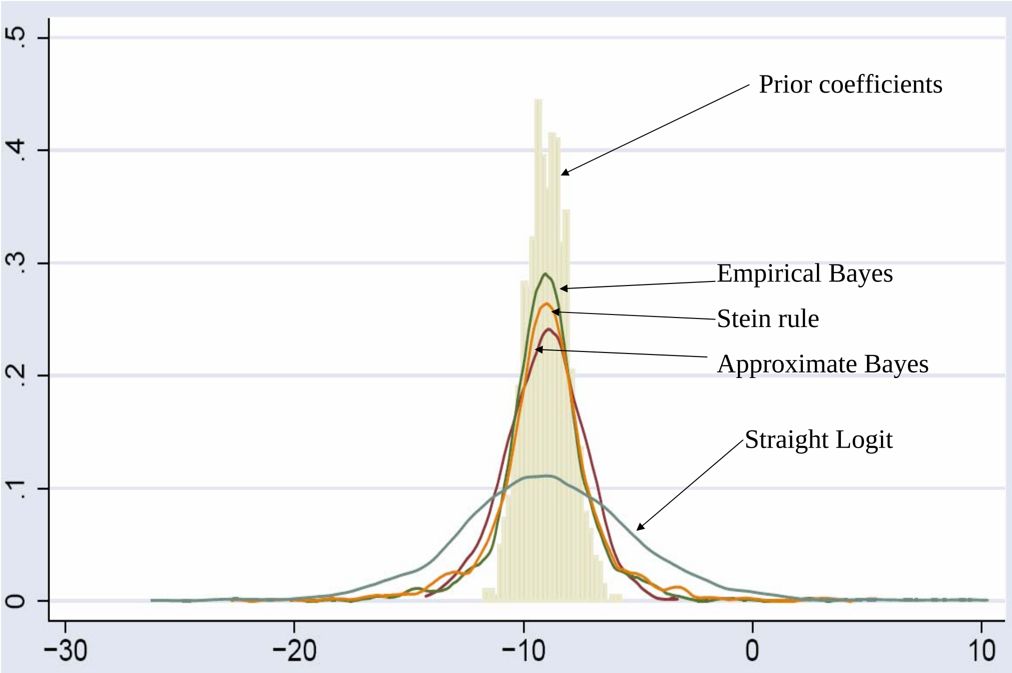

Figure 2: **Frequency Distribution of Simulated Coefficients of** $EBIT/TA$**. S=10%**

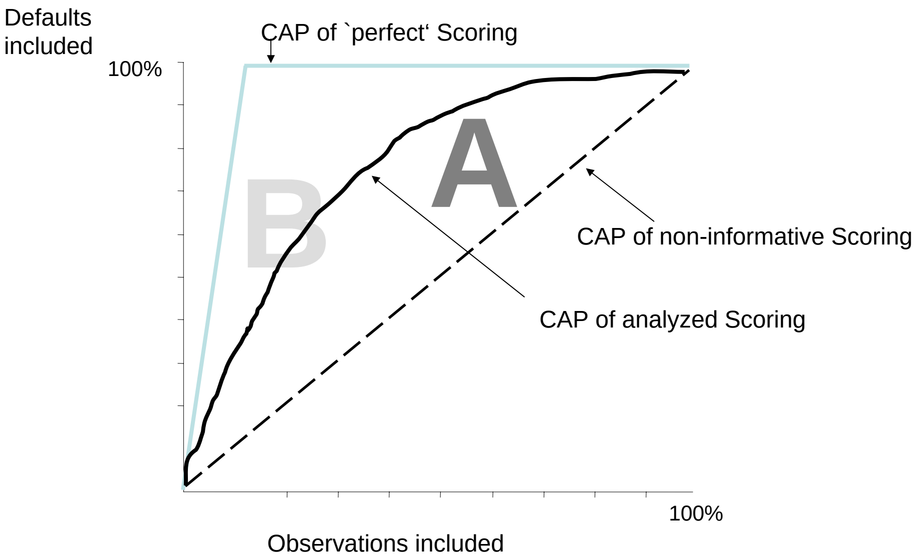

Figure 3: **Illustration of derivation of the accuracy ratio** (A/B).

	WC/TA	RE/TA	EBIT/TA	MV/TL	S/TA	BV/TL
Mean	0.208	0.262	0.083	5.408	1.096	1.675
Median	0.176	0.262	0.087	2.035	0.928	0.848
SD	0.214	0.325	0.110	9.868	0.727	2.490
5% Percentile	-0.072	-0.179	-0.096	0.331	0.287	0.170
95% Percentile	0.616	0.728	0.241	23.639	2.623	5.767

Table 1: **Summary Statistics for Explanatory Variables**

	S=5% accuracy ratio	S=10% accuracy ratio
Mean	80.55	80.59
Median	80.56	80.66
SD	2.10	2.07
5% Percentile	76.95	77.12
95% Percentile	83.81	83.84

Table 2: **Accuracy Ratio for Prior Coefficients (in %).** This table gives descriptive statistics of the accuracy ratio using the Prior Coefficient vector in the simulation. The simulation was repeated $N = 1,000$ times with randomly drawn subsamples of size S (see section 2 for a detailed description of the simulation set-up.) S is the size of the small data set as percentage of the expanded data set, which contains 29,500 firm-years. SD refers to the standard deviation.

	S=5%				S=10%			
	ABE	EBE	SRE	SLE	ABE	EBE	SRE	SLE
A. accuracy ratio of Bayesian estimators & straight logit								
Mean	79.13	78.12	77.49	73.41	79.59	79.28	78.98	76.82
Median	79.31	79.19	79.01	75.93	79.59	79.62	79.45	77.77
SD	1.96	4.65	6.81	8.79	1.27	2.09	2.46	3.75
5% Percentile	75.75	69.99	68.77	59.79	77.42	76.11	74.98	69.96
95% Percentile	81.95	81.98	81.98	80.79	81.66	81.72	81.67	80.85
B. accuracy ratio of Bayesian estimators relative to straight logit								
Mean	5.71	4.71	4.07		2.78	2.47	2.16	
Median	3.01	2.86	2.86		1.64	1.71	1.68	
SD	8.32	6.75	3.85		3.34	2.57	1.90	
5% Percentile	0.23	0.30	0.30		0.15	0.16	0.12	
95% Percentile	17.87	14.81	12.51		9.18	6.96	5.77	
Count	99.7%	98.9%	98.7%		99.4%	98.4%	98.2%	

Table 3: **Simulated Accuracy Ratio in the Base Case (in %)**. Panel A of this table gives descriptive statistics of the accuracy ratio in the simulation. Higher positive values indicate better accuracy. Panel B compares the accuracy ratio of the three Bayesian estimators relative to the straight logit estimation. ABE denotes the approximate Bayes estimator, SRE the Stein Rule estimator, EBE the empirical Bayes estimator and SLE refers to a straight (i.e. standard) logit estimator (see section 3 for details). The simulation was repeated $N = 1,000$ times with randomly drawn subsamples of size S (see section 2 for a detailed description of the simulation set-up.) S is the size of the small data set as percentage of the expanded data set, which contains 29,500 firm-years. SD refers to the standard deviation. 'Count' gives the percentage of simulation repetitions in which one estimator performed better than the other.

	S=5%				S=10%			
	ABE	EBE	SRE	SLE	ABE	EBE	SRE	SLE
A. accuracy ratio of Bayesian estimators & straight logit								
Mean	77.43	77.18	76.50	73.41	78.97	78.98	78.57	76.82
Median	78.66	78.91	78.63	75.93	79.28	79.49	79.25	77.77
SD	5.08	5.92	7.59	8.79	2.22	2.49	2.95	3.75
5% Percentile	66.95	65.68	64.02	59.79	75.52	74.94	73.42	69.96
95% Percentile	81.75	81.88	81.82	80.79	81.51	81.65	81.56	80.85
B. accuracy ratio of Bayesian estimators relative to straight logit								
Mean	4.02	3.77	3.09		2.78	2.47	2.16	
Median	2.40	2.62	2.28		1.64	1.71	1.68	
SD	6.55	5.58	2.88		3.34	2.57	1.90	
5% Percentile	0.09	-0.04	0.07		0.15	0.16	0.12	
95% Percentile	12.59	10.45	8.91		9.18	6.96	5.77	
Count	96.7%	94.8%	95.7%		99.4%	98.4%	98.2%	

Table 4: **Simulated Accuracy Ratio if EBIT/TA is missing in the prior vector (in %).** Panel A of this table gives descriptive statistics of the accuracy ratio in the simulation. Higher positive values indicate better accuracy. Panel B compares the accuracy ratio of the three Bayesian estimators relative to the straight logit estimation. ABE denotes the approximate Bayes estimator, SRE the Stein Rule estimator, EBE the empirical Bayes estimator and SLE refers to a straight (i.e. standard) logit estimator (see section 3 for details). The simulation was repeated $N = 1,000$ times with randomly drawn subsamples of size S (see section 2 for a detailed description of the simulation set-up.) S is the size of the small data set as percentage of the expanded data set, which contains 29,500 firm-years. SD refers to the standard deviation. 'Count' gives the percentage of simulation repetitions in which one estimator performed better than the other.

	S=5%				S=10%			
	ABE	EBE	SRE	SLE	ABE	EBE	SRE	SLE
A. accuracy ratio of Bayesian estimators & straight logit								
Mean	78.19	77.78	77.03	73.41	79.18	79.15	78.72	76.82
Median	78.64	78.91	78.48	75.93	79.27	79.48	79.19	77.77
SD	4.08	4.98	6.91	8.79	1.50	2.09	2.50	3.75
5% Percentile	73.59	69.34	68.30	59.79	76.75	76.12	74.20	69.96
95% Percentile	81.67	81.75	81.63	80.79	81.52	81.65	81.51	80.85
B. accuracy ratio of Bayesian estimators relative to straight logit								
Mean	4.77	4.37	3.62		2.36	2.34	1.90	
Median	2.43	2.85	2.40		1.30	1.66	1.41	
SD	7.47	6.24	3.73		3.03	2.49	1.85	
5% Percentile	0.07	-0.10	-0.02		0.06	-0.05	0.00	
95% Percentile	15.85	13.56	12.11		7.82	6.87	5.53	
Count	96.2%	93.9%	94.9%		96.7%	94.2%	95.2%	

Table 5: **Simulated Accuracy Ratio if MV/TL is missing in the prior vector (in %).** Panel A of this table gives descriptive statistics of the accuracy ratio in the simulation. Higher positive values indicate better accuracy. Panel B compares the accuracy ratio of the three Bayesian estimators relative to the straight logit estimation. ABE denotes the approximate Bayes estimator, SRE the Stein Rule estimator, EBE the empirical Bayes estimator and SLE refers to a straight (i.e. standard) logit estimator (see section 3 for details). The simulation was repeated $N = 1,000$ times with randomly drawn subsamples of size S (see section 2 for a detailed description of the simulation set-up.) S is the size of the small data set as percentage of the expanded data set, which contains 29,500 firm-years. SD refers to the standard deviation. 'Count' gives the percentage of simulation repetitions in which one estimator performed better than the other.

	S=5%				S=10%			
	ABE	EBE	SRE	SLE	ABE	EBE	SRE	SLE
A. accuracy ratio of Bayesian estimators & straight logit								
Mean	72.81	70.75	69.97	63.42	72.94	72.06	71.71	68.46
Median	72.88	72.37	72.24	67.24	73.04	72.66	72.41	69.96
SD	2.30	7.25	8.32	12.56	1.61	3.48	3.82	5.81
5% Percentile	68.75	59.28	56.33	37.76	70.12	66.93	65.03	57.72
95% Percentile	76.20	76.18	76.12	74.32	75.41	75.38	75.32	74.29
B. accuracy ratio of Bayesian estimators relative to straight logit								
Mean	9.40	7.33	6.56		4.48	3.60	3.25	
Median	5.06	4.37	4.21		2.69	2.53	2.52	
SD	12.15	8.99	7.09		5.28	3.56	2.99	
5% Percentile	0.31	0.27	0.26		0.16	0.15	0.13	
95% Percentile	34.89	23.93	20.46		14.41	10.75	9.21	
Count	97.7%	97.2%	96.7%		97.3%	96.8%	96.7%	

Table 6: **Simulated Accuracy Ratio if MV is replaced with BV (in %).** Panel A of this table gives descriptive statistics of the accuracy ratio in the simulation. Higher positive values indicate better accuracy. Panel B compares the accuracy ratio of the three Bayesian estimators relative to the straight logit estimation. ABE denotes the approximate Bayes estimator, SRE the Stein Rule estimator, EBE the empirical Bayes estimator and SLE refers to a straight (i.e. standard) logit estimator (see section 3 for details). The simulation was repeated $N = 1,000$ times with randomly drawn subsamples of size S (see section 2 for a detailed description of the simulation set-up.) S is the size of the small data set as percentage of the expanded data set, which contains 29,500 firm-years. SD refers to the standard deviation. 'Count' gives the percentage of simulation repetitions in which one estimator performed better than the other.

Appendix

	S=5%				S=10%			
	ABE	EBE	SRE	SLE	ABE	EBE	SRE	SLE
A. Brier Score of Bayesian estimators & straight logit								
Mean	0.430	0.429	0.476	0.490	0.424	0.424	0.425	0.431
Median	0.429	0.427	0.428	0.437	0.424	0.424	0.425	0.431
SD	0.034	0.034	1.050	1.125	0.025	0.025	0.025	0.026
5% Percentile	0.376	0.376	0.376	0.382	0.385	0.385	0.385	0.390
95% Percentile	0.487	0.486	0.488	0.501	0.467	0.467	0.468	0.477
B. Brier Score of Bayesian estimators relative to straight logit								
Mean	-0.060	- 0.065	-0.018		-0.007	-0.007	-0.007	
Median	-0.005	- 0.010	-0.010		-0.005	-0.006	-0.005	
SD	1.125	1.125	0.146		0.006	0.006	0.005	
5% Percentile	-0.001	- 0.003	-0.002		-0.001	-0.001	-0.001	
95% Percentile	- 0.031	-0.039	-0.037		-0.018	-0.019	-0.018	
Count	97.8%	99.9%	99.8%		99.6%	98.5%	98.3%	

Table 7: **Simulated Brier Score in the Base Case (in %).** Panel A of this table gives descriptive statistics of the Brier score in the simulation. Lower values indicate better accuracy. Panel B compares the Brier score of the three Bayesian estimators relative to the straight logit estimation. ABE denotes the approximate Bayes estimator, SRE the Stein Rule estimator, EBE the empirical Bayes estimator and SLE refers to a straight (i.e. standard) logit estimator (see section 3 for details). The simulation was repeated $N = 1,000$ times with randomly drawn subsamples of size S (see section 2 for a detailed description of the simulation set-up.) S is the size of the small data set as percentage of the expanded data set, which contains 29,500 firm-years. SD refers to the standard deviation. 'Count' gives the percentage of simulation repetitions in which one estimator performed better than the other.

References

Adkins, L. and R. C. Hill (1996), Using prior information in the probit model: Empirical risks of bayes, empirical bayes, and stein estimators, *in* D. A.Berry, K. M.Chaloner and J. K.Geweke, eds, 'Bayesian Analysis in Statistics and Econometrics', Wiley.

Altman, E. I. (1968), 'Financial ratios, discriminant analysis, and the prediction of corporate bankcrupty', *Journal of Finance* **23**, 589–609.

Banque de France (2001), Recent developments in the Banque de France's scoring method. Banque de France Bulletin Digest, No. 93.

Barnett, V. and T. Lewis (1994), *Outliers in statistical data*, 3rd edn, Wiley.

Basel Committee on Banking Supervision (2003), 'Overview of the new basel capital accord'.

Beaver, W. (1968), 'Alternative accounting measures as predictors of failure', *The Accounting Review* pp. 113–122.

Brier, G. W. (1950), 'Verification of forecasts expressed in terms of probability', *Monthly Weather Review* **78**, 1–3.

Chava, S. and R. A. Jarrow (2004), 'Bankruptcy prediction with industry effects', *Review of Finance* **8**(4), 537 – 569.

Engelmann, B., E. Hayden and D. Tasche (2003), 'Testing rating accuracy', *Risk* **16**, 82–86.

Frerichs, H. and M. Wahrenburg (2003), Evaluating internal credit rating systems depending on bank size. Working Paper Series: Finance & Accounting, J.W. Goethe-University Frankfurt/Main. No. 115.

Greene, W. H. (2003), *Econometric Analysis*, Prentice Hall.

Grunert, J., L. Norden and M. Weber (2005), 'The role of non-financial factors in internal credit ratings', *Journal of Banking and Finance* **29**, 509–531.

James, W. and C. Stein (1961), Estimation with quadratic loss, *in* 'Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability', University of California Press, Berkeley, pp. 361–379.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl and T. Lee (1985), *The Theory and Practice of Econometrics*, 2nd edn, Wiley and Sons.

Shumway, T. (2001), 'Forecasting bankruptcy more accurately: A simple hazard model', *Journal of Business* **74**, 101–124.

Sobehart, J. R. and S. C. Keenan (2001), 'Measuring default accurately', *Risk (Credit Risk Special Report)* **14**, 31–33.

Stein, R. M. (2005), 'The relationship between default prediction and lending profits. integrating ROC analysis and loan pricing.', *Journal of Banking & Finance* **29**, 1213–1236.

Stein, R. M. and F. Jordao (2003), 'What is a more powerful model worth?', *Moody´s KMV* .

Zellner, A. and P. E. Rossi (1984), 'Bayesian analysis of dichotomous quantal response model', *Journal of Econometrics* **25**, 365–393.