

Paper 45-2010

Evaluation of methods to determine optimal cutpoints for predicting mortgage default

Valentin Todorov, Assurant Specialty Property, Atlanta, GA
Doug Thompson, Assurant Health, Milwaukee, WI

Abstract

One of the lessons from the recent wave of mortgage defaults is that it is very important to know the level of risk associated with mortgages held in a portfolio. In the banking world, lending risk is usually defined in terms of the likelihood that a borrower will default on a mortgage. To estimate the probability of default, banks utilize various risk scoring techniques, which output a score. Depending on a predetermined threshold “cutpoint” for the score, loans are priced differently. Loans with scores under that cutpoint are assigned higher interest rates compared to loans with scores above the cutpoint. If there is no experience available to choose an optimal cutpoint, statistical criteria can be used. One method for determining the cutpoint, the Kolmogorov-Smirnov goodness-of-fit test, has become the industry standard. Other less commonly used methods are decision trees/rules and a series of chi-square tests. In this paper we evaluate and compare these methods for estimating an optimal cutpoint. The methods illustrated here are applicable in many fields some of which are finance, biostatistics, and insurance.

Introduction

Risk models that determine borrowers’ credit quality are widely used in financial institutions. The models usually estimate the ability of a person to repay debt and the outcome influences decisions for extending credit or not. Even if credit is extended, banks may charge different interest rates depending on the level of a loan’s estimated risk. For example, low risk loans, referred to as Prime, usually have lower interest rates compared to high risk loans, referred to as Subprime.

With regard to loans, the decisions that banks make are usually binary, e.g. extend credit to an applicant (Yes/No), loan will default (Yes/No), etc. These types of decisions can be modeled using logistic regression. The output of logistic regression is a score that represents the probability that the modeled event will occur. Using the score, loans can be classified as low or high risk, relative to a predetermined score threshold, known as a “score cutpoint”.

If sufficient data is available, analysts can choose a score cutpoint that maximizes or minimizes a target outcome – for example, they may choose a score cutpoint that maximizes revenue or minimizes losses resulting from defaults. However, sometimes the data available is insufficient to enable choosing a score cutpoint on this basis. For example, when starting a new line of business, there may be no historical data available to choose a score cutpoint based on past experience with revenue, profit, etc. To begin the business, one must choose an initial score cutpoint based on whatever data is available. If the analyst has insufficient data available to select a score cutpoint that maximizes or minimizes the desired outcome, one option is to use statistical criteria to maximize the strength of association between a score cutpoint and a binary target that roughly approximates the desired outcome. In other words, set a cutpoint such that the resulting 2 x 2 table (rows: observation above cutpoint vs. not above cutpoint, columns: observation in one category of the binary dependent variable vs. in the other category) shows the

strongest possible association between the row and column variables. (Exactly what this means is described in more detail below.) This paper describes and illustrates methods for determining an optimal score cutpoint based on statistical criteria. Specifically, we will compare three methods for finding optimal cutpoint for a score – Kolmogorov – Smirnov goodness-of-fit test, decision trees/rules, and chi-square tests.

Mortgage default model

Since the early 2007 the US housing market has suffered severely as a result of a sharp increase in mortgage defaults and foreclosures. Foreclosure is the process that banks initiate to repossess properties whose owners have stopped paying the mortgage on the loan obtained to purchase that property. In many cases, the reason a property goes into foreclosure is because lenders have extended credit to borrowers without due consideration of the financial stability of the credit applicants. For a period of about six years, from 2000 to 2006, housing prices across the whole US were steadily increasing. The economy was booming and people overextended their finances to purchase homes. Borrowers with unstable financial situation were classified as Subprime and given mortgages with high interest rates, the payment of which caused widespread defaults a few years later.

The model score on which we demonstrate the comparison of the cutpoint techniques is the probability that a loan will go into foreclosure. Various factors can influence the decision to default on a mortgage, but among the main ones are borrowers' personal savings and income, amount of the monthly mortgage payment, type of interest rate (fixed or variable) and the employment situation (employed or unemployed). Using a publicly available dataset we modeled the likelihood that a property will go into foreclosure testing some of the variables discussed.

The data used is the *2009 Housing, Mortgage Distress and Wealth* dataset from the Panel Study of Income Dynamics (PSID). This is a balanced panel of 7,922 families that were active in 2007 and 2009. Families participating in the survey were asked questions about their housing situation, whether they own a house or rent, and their personal savings and income. Homeowners were asked additional questions about the type of their mortgage, interest rate, and if they are in foreclosure, which is our variable of interest.

The model was developed using a logistic regression, implemented using PROC LOGISTIC in SAS ®. The dependent variable was coded 1 = property in foreclosure vs. 0 = property not in foreclosure.

The logit is calculated as

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta'x$$

where $\pi = P(Y = 1 | x)$ is the probability modeled, α is the intercept and β is a vector of the coefficients for the independent variables.

To calculate the model score from the logit function we use the transformation below. For presentation purposes the score is multiplied by 1000.

$$\text{score} = \left(\frac{1}{1 + e^{-\text{logit}(\pi)}}\right) * 1000$$

To find the variables that have the strongest association with the outcome, we used variable reduction techniques, which narrowed the predictor set to five covariates which together maximized predictive performance of the model. The logistic regression model in SAS for the selected variables is presented below.

```

%let depvar      = mortg_dq;
%let weight      = sample_weight;
%let predictors  = bankacct_total_amt
                  fixed_i_rate
                  interest_rate
                  mtg1_years_to_mat
                  own_ira_stock;

proc logistic data=compsamp descending namelen=100;
weight &weight.;
model &depvar. = &predictors.;
run;

```

The final model exhibited good predictive performance. The c-statistic on the holdout sample was 0.86, demonstrating a good fit with the data. All of the selected predictors were highly significant, excluding the fixed interest rate binary variable (*fixed_i_rate*). This variable was only marginally significant (p-value=0.032), but we decided to keep it in the model, due to its known impact on the decision of borrowers to default. Borrowers with a variable interest rate are usually more likely to default, because their rates are continuously adjusted upward by lenders, which increases the monthly mortgage payments due.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.9066	0.7628	41.3786	<.0001
bankacct_total_amt	1	-0.00009	0.000026	12.0736	0.0005
fixed_i_rate	1	-0.6650	0.3105	4.5856	0.0322
interest_rate	1	0.3060	0.0677	20.4402	<.0001
mtg1_years_to_mat	1	0.0656	0.0163	16.2223	<.0001
own_ira_stock	1	-1.7803	0.4370	16.5938	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
bankacct_total_amt	1.000	1.000	1.000
fixed_i_rate	0.514	0.280	0.945
interest_rate	1.358	1.189	1.551
mtg1_years_to_mat	1.068	1.034	1.102
own_ira_stock	0.169	0.072	0.397

Comparison of results to determine optimal score “cutpoint”

Kolmogorov-Smirnov goodness-of-fit test

In the private sector, the Kolmogorov-Smirnov goodness-of-fit test is a popular method to determine appropriate score cutoff points. What makes it so popular is its relative simplicity, compared to other statistical tests, which is important when results are presented to decision makers who may not have training in statistics. Another feature that makes the test appealing is that the empirical distribution functions of both groups tested can be graphed and presented to stakeholders.

The Kolmogorov-Smirnov test is based on the empirical distribution functions of the modeled scores across two groups – loans in foreclosure and not in foreclosure. Under the null hypothesis of the test, there is no difference in the distribution of scores in both groups, and under the alternative hypothesis the distributions are different. In SAS the test is performed using the PROC NPAR1WAY with the option EDF. The procedure computes the Kolmogorov-Smirnov test statistic D , which is distance between both EDFs for every value of the tested score. The score at which the statistic D is maximized is the score that can best discriminate between both groups.

The KS test statistic D is calculated as

$$D = \max | F_1(x_i) - F_2(x_i) |,$$

where $F_1(x_i)$ and $F_2(x_i)$ are the EDFs of both groups of loans.

To test on the empirical distribution functions it is important to specify the EDF option for PROC NPAR1WAY.

```
%macro kstest(inset=, depvar=, outset=, scoreVar=);  
proc npar1way data=&inset. EDF;  
class &depvar.;  
var &scoreVar.;  
output out=&outset. EDF;  
run;  
%mend;
```

```
%kstest (inset=train2, depvar=mortg_dq, outset=out, scorevar=score);
```

The Kolmogorov-Smirnov goodness-of-fit test was performed on a random sample of 200 observations, which is about 7% of the total file. We used a sample, because we found that the third method presented in this paper (Chi-square tests) is not suitable for datasets with more than 200 observations. Based on the p-value for the test we can reject the null hypothesis that the EDFs of both groups are the same.

Kolmogorov-Smirnov Two-Sample Test (Asymptotic)

KS	0.223128	D	0.608631
KSa	3.155504	Pr > KSa	<.0001

At the point where the distance between the EDFs is at a maximum ($D=0.609$), we find our suggested cutpoint. The suggested cutoff point for our data is 88, which is found at the bottom of the table below. All loans with a score above the cutpoint are classified as High Risk and the loans below the cutpoint are classified as Low Risk.

Kolmogorov-Smirnov Test for Variable score
Classified by Variable mortg_dq

mortg_dq	N	EDF at Maximum	Deviation from Mean at Maximum
1	32	0.218750	-2.892067
0	168	0.827381	1.262201
Total	200	0.730000	

Maximum Deviation Occurred at Observation
155

Value of score at Maximum = 87.957772

Decision trees/rules

Another approach to select an optimal score cutpoint is to use decision trees or decision rules. Using this method, one begins with all observations combined in a single category. Then, one splits the observations into two categories, based on whether or not the observations are above the score cutpoint. A score cutpoint is chosen that best separates the observations into two categories, with respect to the “purity” of these categories on the target variable – specifically, one category should have a maximal proportion of one level of the outcome (e.g., percent in foreclosure) while the other category should have a maximal proportion of the other outcome level (e.g., percent not in foreclosure). These algorithms mine through all possible cutpoints (using shortcuts when possible, to be efficient) to identify the optimal cutpoint. There are many variants of decision tree/rule algorithms and they may yield different score cutpoints.

SAS Enterprise Miner is an excellent tool for decision tree/rule analyses. However, some companies do not purchase SAS Enterprise Miner for cost reasons. Alternatives for decision tree/rule analyses include freeware programs such as Weka, R or RapidMiner. In this paper, we identified score cutpoints using algorithms available in Weka (decisions tree/rule algorithms available in Weka are described by Witten & Frank, 2005). To do this, we created a SAS dataset including the target outcome (foreclosure vs. not) and the model score described above. The dataset analyzed using Weka included only observations in the training sample. This dataset was exported as a CSV file and then imported into Weka. Different decision tree/rule algorithms in Weka were used to identify a set of optimal score cutpoints – specifically, we tried the JRIP, RIDOR, J48 and RANDOMTREE algorithms in Weka. These cutpoints were further evaluated on the holdout sample in SAS, to ensure that they would generalize. Score cutpoints that looked especially promising on the holdout, namely cutpoints of 134 and 42, were retained for further consideration (see *Comparing the cutpoints* below).

Chi-square tests (Williams et al.)

The final method for finding an optimal cutpoint presented in this paper uses a macro developed by researchers at the Mayo Clinic (Williams et al.). At a high level, the macro performs a series

of chi-square tests for all potential cutpoints in search for the one with the minimum p-value and the maximum odds ratio.

The macro performs the selection in two steps. First, it calculates p-values and odds ratios for potential cutpoints and then it ranks the selected cutpoints by their ability to split the data in two groups.

At the first step the macro performs a series of two-sample tests for all possible dichotomizations of the model score. For each unique model score the macro performs a chi-square test to determine p-value. The best cutpoint will have the lowest possible p-value. Since the methodology performs multiple tests in search for the minimum p-value, it is possible that the significance of the final score is overestimated. To account for this problem, the macro calculates an adjusted p-value for each cutpoint using methodology suggested by Miller and Seigmund (1982). Besides, the p-value the macro calculates also the odds ratios. The cutpoints with the highest odds ratios are potential candidates to use for dichotomization.

After calculating the p-values and odds ratios for all possible cutpoints the macro selects the ten scores with the lowest p-values and the highest odds ratios. Readers interested in more details should refer to Williams et al (2006).

The parameters of the macro can be changed depending on analysts' needs. The parameters we used in our analysis are:

```
%cutpoint (dvar=score,
           endpoint=mortg_dq,
           data=fullc,
           trunc=round,
           type=integers,
           range=ninety,
           fe=on,
           plot=gplot,
           ngroups=2,
           padjust=miller,
           zoom=no);
```

The top ten cutpoints suggested are presented in the table. The cutpoint with the optimal combination of lowest p-value and highest odds ratio is 88.

Suggested Cutpoints	p-value	Adjusted p-value	Odds Ratio	p-value Score	Odds Ratio Score	Total Score
88	1.18E-12	1.78E-10	17.12	10	10	20
87	5.56E-12	7.88E-10	15.78	9	9	18
86	1.16E-11	1.59E-9	15.18	8	7	15
90	1.18E-11	1.62E-9	14.38	7	5	12
80	3.83E-11	5.00E-9	15.34	3	8	11
95	1.18E-11	1.62E-9	14.38	6	4	10
85	2.36E-11	3.14E-9	14.61	4	6	10
105	1.95E-11	2.62E-9	13.20	5	3	8
101	4.49E-11	5.82E-9	12.58	2	2	4
102	4.49E-11	5.82E-9	12.58	1	1	2

These cutpoints were calculated using a random sample of 200 observations, which is about 7% from the complete file. We had to use a small sample, because the p-values calculated from the full dataset are less than 1.9E-16, at which level the comparisons are meaningless. After testing the macro on different sample sizes we concluded that the macro provides consistent results only for small datasets with no more than 200 observations.

Comparing the cutpoints

Above, we described and illustrated three methods to identify optimal score cutpoints based on statistical criteria. We identified score cutpoints based on the Kolmogorov-Smirnov statistic, decision trees/rules and a chi-square test method described by Williams et al.

Different methods can yield different score cutpoints. In choosing an optimal cutpoint, it may make sense to take the cutpoints obtained from different methods and compare them with respect to effect size measures. In this context, effect size is a measure of the strength of association between the binary variable resulting from the score cutpoint (i.e., 1 = above the score cutpoint vs. 0 = not above the score cutpoint) and the binary outcome (i.e., 1 = loan in foreclosure vs. 0 = loan not in foreclosure). Essentially, we are attempting to quantify the strength of association in a 2 x 2 table, where the row dimension represents above vs. below the score cutpoint and the column dimension represents loans in foreclosure vs. not in foreclosure. If the strength of association is maximized, this means that the score cutpoint that was chosen yielded maximum separation of the observations with respect to the target outcome (foreclosure vs. not in foreclosure). In other words, the score cutpoint gives us the maximum ability to put the “good” loans in one category and the “bad” loans in the other category.

There are different ways to estimate strength of association (effect size) in 2 x 2 tables, as described by Grissom and Kim (2005). Two useful measures are relative risk and the phi coefficient. Relative risk is simply percent target among observations with scores above the cutpoint, divided by percent target among observations with scores not above the cutpoint (e.g., % in foreclosure among observations above the cutpoint / % in foreclosure among observations not above the cutpoint). Greater relative risk indicates greater relative difference between the groups with respect to the outcome variable. For 2 x 2 tables, relative risk may be a better measure of effect size than the odds ratio (see Kraemer, 2006, for a discussion of limitations of the odds ratio). The phi coefficient can be thought of as a correlation coefficient for 2 x 2 tables. It is based on the Pearson chi-square statistic. It ranges from -1 to 1 and greater departures from 0 point to greater strength of association, similar to other common correlation coefficients.

To choose an optimal score cutpoint among the cutpoints yielded by various methods (which might differ in some situations), one approach is to compare different cutpoints with respect to relative risk and phi. The best score cutpoint will maximize both relative risk and phi. A table showing these measures for different score cutpoints can be produced using the SAS macro %EVALUATE_SCORE_CUTPOINTS shown below. With minor modifications, this macro could be used to evaluate score cutpoints in almost any dataset.

```

%macro evaluate_score_cutpoints;
* List of cutpoints to be considered;
%let cuts=
88
87
86
42
134;

* Analyses run separately for each cutpoint;
%do i=1 %to 5;
%let curr_cut = %scan(&cuts,&i);

* Define cutpoint indicator as 1 = score > cutpoint, 0 otherwise;
data hold;
set holdout;
cut&curr_cut=(score>&curr_cut);
run;

* 2 x 2 analyses of cutpoint indicator x dependent variable (here,
mortg_dq).;
* Use PROC FREQ to calculate row percentages, relative risk and phi.;
proc freq data=hold;
weight holdwgt;
tables cut&curr_cut*mortg_dq / measures chisq;
ods output RelativeRisks=RelativeRisks CrossTabFreqs=CrossTabFreqs
ChiSq=ChiSq;
run;

* The next 3 data steps gets and formats selected results from ODS OUTPUT;
* tables generated by PROC FREQ.;
data rr2;
set RelativeRisks;
keep relrisk;
if StudyType='Cohort (Col2 Risk)' then do;
relrisk = left(compress( put((1/value),5.3) ));
output;
end;
run;

data phi;
set ChiSq;
keep phi;
if Statistic='Phi Coefficient' then do;
phi = left(compress( put((value),5.3) ));
output;
end;
run;

data above_cut(keep=pct_mortg_dq_above_cut)
not_above_cut(keep=pct_mortg_dq_not_above_cut);
set crosstabfreqs;
if cut&curr_cut=0 and mortg_dq=1 then do;
pct_mortg_dq_not_above_cut=left(compress( put(rowpercent,5.2) || "%" ));
output not_above_cut;
end;
if cut&curr_cut=1 and mortg_dq=1 then do;
pct_mortg_dq_above_cut=left(compress( put(rowpercent,5.2) || "%" ));
output above_cut;
end;

```



```

run;

* Combine the measures in a single table where the first column (cutpt);
* indicates the score cutpoint.;
data allmeasures;
retain cutpt;
merge above_cut not_above_cut rr2 phi;
cutpt=&curr_cut;
run;

* Combine the tables for all cutpoints being considered.;
%if &i=1 %then %do;
  data cutpoint_eval_table;
  set allmeasures;
  run;
%end;
%else %do;
  data cutpoint_eval_table;
  set cutpoint_eval_table allmeasures;
  run;
%end;
%end;
%mend evaluate_score_cutpoints;
%evaluate_score_cutpoints;

```

The table below shows the results produced by %EVALUATE_SCORE_CUTPOINTS. The SAS table produced by the macro was exported to Excel and the column headers were relabeled and formatted.

Score cutpoint	% mortgage defaults for cases above cutpoint	% mortgage defaults for cases not above cutpoint	Relative risk	Phi
88	24.24%	3.31%	7.313	0.316
87	23.08%	3.35%	6.879	0.303
86	22.71%	3.37%	6.744	0.299
42	15.12%	2.85%	5.297	0.227
134	33.33%	4.37%	7.623	0.335

In these data, one of the score cutpoints selected via tree/rule analysis (= 134) exhibited maximum relative risk as well as maximum phi. However, the score cutpoint of 88 chosen by both the Kolmogorv-Smirnov statistics and the Williams et al method were a close second and for practical purposes performed about as well as the cutpoint of 134. Given that 134 and 88 are both viable options, other business considerations might be taken into account, for example the fact that more adverse decisions would be made if the 134 cutpoint was chosen, possibly resulting in customer ill-will and backlash. This might be one reason to choose 88 over 134, despite the fact that 134 had slightly greater relative risk and phi.

Conclusion

In many situations, a predictive model score needs to be cut at some point (the “score cutpoint”) to determine a binary decision, e.g., extend a loan with a Prime vs. Subprime interest rate. An important consideration is how to select the score cutpoint. If sufficient data are available, one may choose a score cutpoint that maximizes or minimizes the target outcome, for example, maximize revenue, minimize losses resulting from defaults. Refer to Siddiqi (2006) for guidance on how to choose an optimal cutpoint on this basis. However, sometimes sufficient data are not available to choose an optimal score cutpoint based on the ultimate target outcome. For example, when starting up a new line of business, there may not be enough data to choose a cutpoint based on the ultimate target outcome such as revenue maximization or dollar loss minimization. In such situations, one may need to use whatever data are available (for example, publically available data) and choose a cutpoint that maximizes separation with respect to a proxy for the ultimate outcome of interest (for example, mortgage default vs. no default). In such situations, one may choose a cutpoint that maximizes ability to separate observations into two categories with respect to the binary, proxy outcome. This paper described and illustrated three methods to determine an optimal score cutpoint in such situations, based on statistical criteria, namely the Kolmogorov-Smirnov statistic, decision trees/rules and a chi-square test method described by Williams et al.

We found that all methods pointed to cutpoints that performed similarly with respect to two effect size measures, relative risk and phi. Relative risk and phi indicate the ability of the binary indicator resulting from a score cutpoint to separate observations into categories that are maximally different with respect to the outcome. Although the methods pointed to similar optimal score cutpoints in the example presented in this paper, this may not be the case in every dataset.

The methods described in this paper can be used in many different fields, for example, insurance (e.g., use a streamlined vs. intensive underwriting process for applicants above vs. not above a score cutpoint), medicine (e.g., patients with lab scores above a certain level receive a particular treatment, whereas those with scores not above that level do not receive the treatment), education (e.g., students with test scores above a certain level are put in an advanced track whereas students with scores not above that level are put in the normal track) and many other fields.

References

Grissom RJ, Kim JJ. (2005). *Effect sizes for research: A broad practical approach*. Mahwah, NJ: Lawrence Erlbaum.

Housing, Mortgage distress and Wealth Data. (2009). *The Panel Study of Income Dynamics*, University of Michigan.

Kraemer HC. (2006). Correlation coefficients in medical research: From product moment correlation to the odds ratio. *Statistical Methods in Medical Research*, 15: 525-545.

Miller R, Seigmund D. (1982). Maximum selected chi-square statistics. *Biometrics*; 38; 1011-1016.

Siddiqi N. (2006). *Credit risk scorecards: Developing and implementing intelligent credit scoring*. New York: Wiley.

Williams et al. (2006). *Finding Optimal Cutpoints for Continuous Covariates with Binary and Time-to-Event Outcomes*, Technical Report #79, Division of Biostatistics, Mayo Clinic.

Witten IH, Frank E. (2005). *Data mining: Practical machine learning tools and techniques*. New York: Elsevier/Morgan Kaufman.

Contact Information

Valentin Todorov
Assurant Specialty Property
260 Interstate North Cir SE
Atlanta, GA 30339
Phone: 770-834-5161
valentin.todorov@assurant.com
www.vatodorov.com

Doug Thompson
Assurant Health
501 W. Michigan St.
Milwaukee, WI 53201
Phone: 414-299-7998
doug.thompson@assurant.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.