Paper AD10-2011

Application of SAS to monitoring loan defaults in consumer credit portfolios

Yevhen Yankovskyy, ATB Financial, Edmonton, Alberta, Canada[1]

Abstract

A current economic downturn has negatively affected borrowers' solvency in most communities across North America. However, some communities with either over-heated property market or high dependency on a few employers suffered more than the average. A geographic default monitoring system described in this paper is designed to analyze and effectively present information on default risk concentration in the neighbouring residential areas. This system consists of the following key elements: a) constructing several funnel plots at reasonable confidence levels, b) classification of the communities into risk groups by location of their default rates at the funnel plots, c) transferring borrowers' classification information onto maps with assigning certain colours corresponding to specific default risk groups. Having the analysis results in hand, portfolio managers will be able to detect areas with either high or low default risks and make qualified decisions on changes in their credit policies.

Introduction

A current economic downturn has negatively affected borrowers' solvency in most of the communities across North America. However, some communities with either over-heated property market or high dependency on a few employers suffered more than the average. The geographical default monitoring system described below is designed to analyze and effectively present information on default risk concentrated in the neighbouring areas. It provides statistically valid ground for classification of the consumers by their risk of default. Entering the classification results on the chart will allow detecting areas with concentration of risky customers and induces managers to make corresponding changes in their credit policies in these areas.

Part 1. Methodology of risk level funnel plots

To assess default risks observed in an area, I suggest using risk level funnel plots. A risk level funnel plot is originated from the normal approximation for a sampling distribution of a ratio (Anderson, Sweeney, Williams. (2005)). A single funnel plot is a chart combining observed sample rates in the areas with the line connecting confidence intervals for the population rate estimate which are estimated for all possible sample sizes at the same confidence level. The funnel plots have proven to be effective and statistically valid tool for comparing the rates in small samples and often used in epidemiology for monitoring outbreaks of the contagious diseases (Dover, 2010).

In banking industry the funnel plots can be applied to the borrowers' default rates (i.e., 30-day, 60-day, and 90-day delinquency rates) in consumer credit portfolios and aimed at detection of areas with either significantly high or low deviation of the observed delinquency rates from the expected delinquency rate given the portfolio size. In practise, a risk analyst faces two issues: a) comparing the default rates from the areas with different portfolio sizes (also called loan volumes), b) monitoring concentration of defaults in the neighbouring areas.

To solve these issues, I offer to apply a geographic monitoring system that consists of a) several funnel plots evaluated at the meaningful confidence levels, b) qualitative classification of the funnel plots results, and c) mapping their results onto geographic maps.

The application of the funnel plots is straightforward in consumer credit portfolios, since these portfolios are quite homogeneous. These portfolios usually consist of standardized loan products sold to many borrowers that are relatively identical in their wealth. In contrast, application of the method commercial portfolios has strong reservations. The commercial portfolios are more heterogeneous and more concentrated, since they often consist of the loans that are customized and sold to businesses of different size, specialization, and areas of operation.

An algorithm for constructing funnel plots

1. Screen available data and define natural geographical borders for the areas with relatively similar population size.

[1] The views and opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of an employer, ATB Financial.

2. Select the areas with sufficient number of the borrowers (at least 30) living in the area. If the neighbouring areas have insufficiently small borrowers' samples, merge them into the bigger entities. Otherwise, exclude the insufficiently small communities from the study.

3. Estimate approximate $(1-\alpha)\%$ confidence intervals (Anderson, Sweeney, Williams, 2005, p. 280) for the population default rates $(\hat{p})$ aggregated for all observed sample sizes (n_i):

$$\hat{p} = \frac{1}{K}\sum_{i=1}^{K}\frac{D_i}{n_i}$$
$$I_{[0;1]}\left(\hat{p} \pm Z_{(1-\alpha/2\,)}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \qquad (1.1)$$
$$n \,\epsilon\, [\min(n_i)\,, \max(n_i)]$$

where

D_i is a number of defaulted loans in all areas of interest i, i= 1, ..., K;

$\hat{p}$ is a overall population default rate estimated by an average default rate in all areas of interest;

$Z_{(1-\alpha/2\,)}$ is a value of the cumulative normal distribution that corresponds to $(1-\alpha/2\,)\%$ probability.

$I_{[0;1]}$ is an indicator function that limits the approximate confidence interval boundaries within a meaningful range.

4. Develop a chart with sample sizes (n_i) with default rates and their confidence intervals on y-axe and all potential borrowers' sample sizes on x-axe. At this chart a reference line can be drawn to mark a portfolio estimate default rate.

5. Connect the confidence interval boundaries for all consecutive sample sizes at the same confidence level to make a funnel plot.

6. Develop several funnel plots at some meaningful confidence level and map them together with observed default rates on the same chart.

Application of colours in mapping risk groups

Table 1. A summary of the default geographic system

Risk level	Corresponding statistic confidence interval boundaries	Color (SAS color)	Suggested credit policy applied in the area
Extremely high	95% and higher	Red	Emergency intervention: stop lending in the area, investigation of the situation
Moderate High	90-95%	Pink	Toughening the credit policy: higher credit limits and down-payment sizes; higher customers' risk score requirements
Normal	10-90%	Green (lime)	No change in credit policy, keep monitoring
Moderate Low	5%-10%	Light blue (cyan)	Loosening the credit policy: lower credit limits, down payment sizes; lower customers' risk score requirements
Extremely Low	5% and lower	Blue	Applying aggressive risk taking credit policy: significantly lower credit limits, down payment sizes, lower customers' risk score requirements

Part 2. An application example

To demonstrate how the geographic default monitoring system works, we apply it to a 90-day delinquency rate and portfolio sizes in consumer credit card portfolio that were randomly generated for the US states (i= 1,…,52) with the following assumptions on distribution of the variables:

90-day delinquency rate$_i$ ~*Beta*(α=0.98, β=97.02)

Loan volume$_i$ ~ 10**Negative Binomial*(p=0.05,k= 2)

Results of running a SAS code are provided below while the code itself is provided in full in the Appendix.

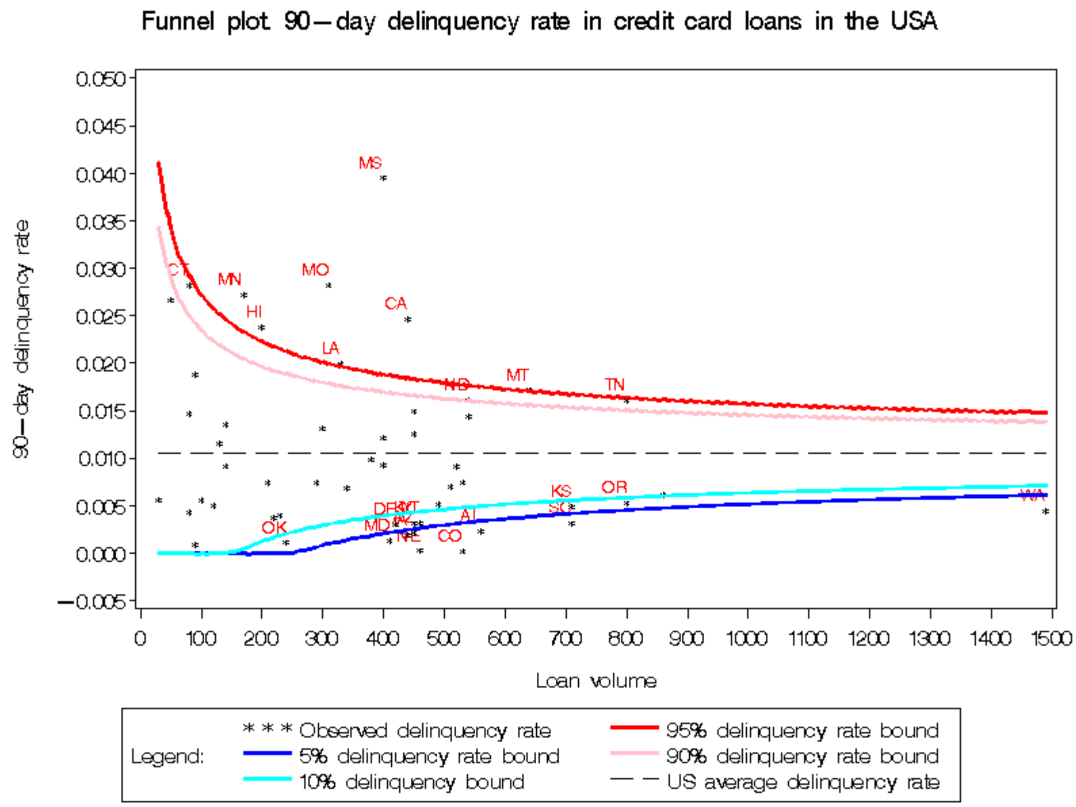


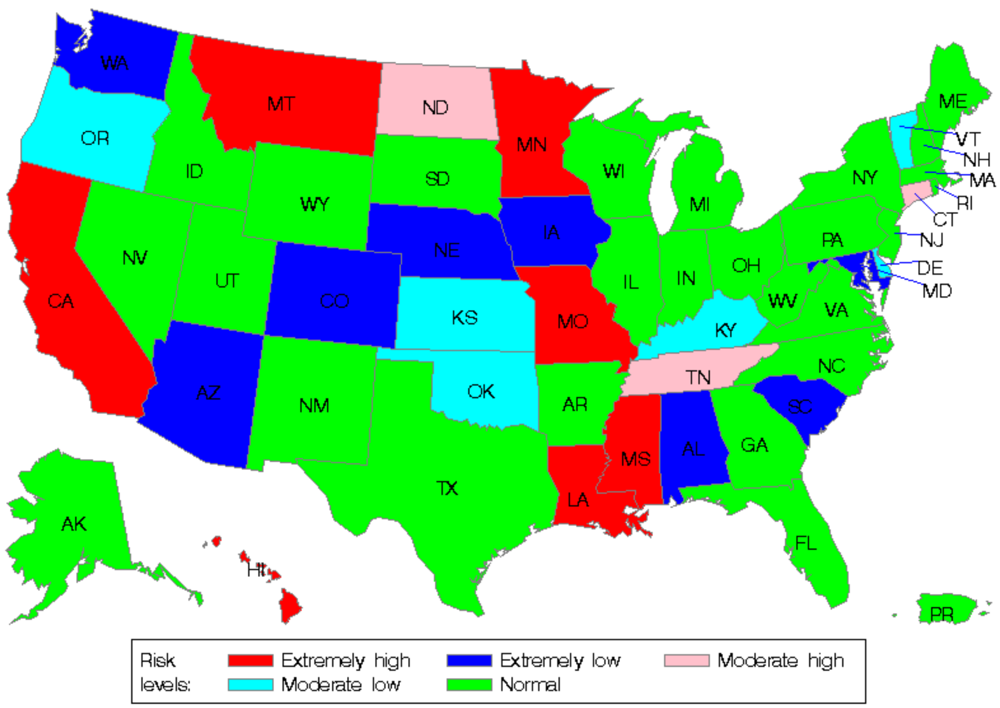
Table. The US states with high delinquency rate

State name abbreviation	Name of State	Loan volume	Observed delinquency rate	90% delinquency rate bound
TN	Tennessee	800	.01612	.01508
MT	Montana	640	.01719	.01562
ND	North Dakota	540	.01615	.01608
CA	California	440	.02467	.01668
MS	Mississippi	400	.03951	.01698
LA	Louisiana	330	.01997	.01764
MO	Missouri	310	.02825	.01787
HI	Hawaii	200	.02377	.01968
MN	Minnesota	170	.02720	.02046
CT	Connecticut	80	.02817	.02504

Table. The US states with low delinquency rate

State name abbreviation	Name of State	Loan volume	Observed delinquency rate	10% delinquency bound
WA	Washington	1490	.00444	.00709
OR	Oregon	800	.00526	.00586
KS	Kansas	710	.00493	.00557
SC	South Carolina	710	.00313	.00557
AL	Alabama	560	.00231	.00496
CO	Colorado	530	.00022	.00480
NE	Nebraska	460	.00026	.00439
VT	Vermont	460	.00319	.00439
AZ	Arizona	450	.00221	.00432
KY	Kentucky	450	.00313	.00432
IA	Iowa	440	.00194	.00425
DE	Delaware	420	.00310	.00411
MD	Maryland	410	.00132	.00403
OK	Oklahoma	240	.00112	.00205

Chart. Classification of credit card portfolios by 90–day delinquency rate in the USA



Conclusion

This paper presents a methodology and SAS code for geographic default monitoring system that offers tools for a) classification of a loan default rates into risk groups using funnel plots, b) mapping the classification results onto geographic maps, c) offering changes in the credit policy depending on default risk level observed in the area. The application of this system will provide a bank analyst or a portfolio manager with necessary information to make qualified decisions regarding their credit policies.

An application of this methodology to randomly generated loan default and volume data for the United States demonstrates capabilities of the system. The developed code uses SAS data maps for the USA, while in reality an analyst will likely have to use more detailed geographic maps (e.g., county level maps) that are not provided by SAS. In this case, an analyst will have either convert non-SAS supported maps into SAS data file or to use other software (e.g., ArcGIS) to put the results estimated in SAS onto a map.

Appendix. Below there is the SAS code discussed in the paper.

```
libname maps 'C:\Program Files\SAS\SASFoundation\9.2\maps\';
libname temp 'H:\My SAS Files\Temp';

/* Resetting SAS global graphical settings to accommodate the charts in the best way*/
options fmtsearch=(sashelp.mapfmts);
goptions reset=global rotate=landscape gunit=pct border cback=white
        colors=(blue green lime lipk cyan red)
        ctext=black ftext=swiss htitle=2.5 htext=2.5;

/* Generating random values for volume and delinquency rates at the state level */
data usmap;
  set maps.us2 (keep = STATE STATECODE STATENAME);
  do i =1 to 52;
                        dlqrt=rand('BETA',0.98,97.02);
                                volume=10*rand('NEGBINOMIAL',0.05,2);
          end;
          drop i;
          label dlqrt='Observed delinquency rate';
          label volume='Loan volume';
          label STATECODE='State name abbreviation';
run;

proc sort data=usmap; by STATECODE; run;

/* Constructing risk level bounds and classifying the states by delinquency rate */

proc sql;
          create table  temp.usmap as
          select *,
          avg(dlqrt) as dlqrt_us label='US average delinquency rate',

          /*90% CI: 5%-95% boundaries:  */
          calculated dlqrt_us+1.645*sqrt(calculated dlqrt_us*(1-calculated dlqrt_us)/volume) as dlqrt_95pct
          label='95% delinquency rate bound',

          case when (calculated dlqrt_us-1.645*sqrt(calculated dlqrt_us*(1-calculated dlqrt_us)/volume))<0 then 0
                    when (calculated dlqrt_us-1.645*sqrt(calculated dlqrt_us*(1-calculated dlqrt_us)/volume))>1 then 1
                    else calculated dlqrt_us-1.645*sqrt(calculated dlqrt_us*(1-calculated dlqrt_us)/volume)
          end as dlqrt_5pct label='5% delinquency rate bound',

          /*80% CI: 10%-90% boundaries: 1.281 */
          calculated dlqrt_us+1.281*sqrt(calculated dlqrt_us*(1-calculated dlqrt_us)/volume) as dlqrt_90pct
          label='90% delinquency rate bound',

          case when (calculated dlqrt_us-1.281*sqrt(calculated dlqrt_us*(1-calculated dlqrt_us)/volume))<0 then 0
                    when (calculated dlqrt_us-1.281*sqrt(calculated dlqrt_us*(1-calculated dlqrt_us)/volume))>1 then 1
                    else calculated dlqrt_us-1.281*sqrt(calculated dlqrt_us*(1-calculated dlqrt_us)/volume)
          end as dlqrt_10pct label='10% delinquency bound',

          /* Classifying in risk levels */
          case when (calculated dlqrt_95pct)<=dlqrt then 'Extremely high'
           when (calculated dlqrt_90pct)<=dlqrt<(calculated dlqrt_95pct) then 'Moderate high'
                    when (calculated dlqrt_10pct)<dlqrt<(calculated dlqrt_90pct) then 'Normal'
                    when (calculated dlqrt_5pct)<dlqrt<=(calculated dlqrt_10pct) then 'Moderate low'
                   else 'Extremely low'
                   end as risk_level
from  usmap;
quit;


/* Constructing Chart 1: Funnel plot */
/* Auxiliary work: idetifying and marking the outliers*/
data temp.outliers;
```

```
            set temp.usmap (keep= STATECODE STATENAME volume dlqrt dlqrt_90pct dlqrt_10pct);
            where dlqrt>=dlqrt_90pct or dlqrt<=dlqrt_10pct;
            format dlqrt 6.5 dlqrt_90pct 6.5 dlqrt_10pct 6.5 flag_print $CHAR15.;
            select;
              when(dlqrt>=dlqrt_90pct) flag_print='Up-outlier';
              when(dlqrt<=dlqrt_10pct) flag_print='Down-outlier';
                      otherwise flag_print='Non-outlier';
  end;
   XSYS = '2'; YSYS = '2';
   X = volume; Y = dlqrt;
   FUNCTION='LABEL';
 function='label';
   text=STATECODE;
   size=1;
   position='1';color='red';
  output;
run;

proc sort data=temp.outliers;
by flag_print descending volume; run;


/* Developing a funnel plot */
proc sort data=temp.usmap;
by volume; run;

   symbol1 v=star i=none h=1 w=2 c=black;                    /* Actual Observations */
   symbol2 v=none i=spline line=1 w=2 c=red;                 /* Extremely high*/
   symbol3 v=none i=spline line=1 w=2 c=blue;                /* Extremely low*/
   symbol4 v=none i=spline line=1 w=2 c=pink;                /* Moderate high */
   symbol5 v=none i=spline line=1 w=2 c=cyan;                /* Moderate low: */
   symbol6 v=none i=spline line=1 w=2 c=lime;                /* Normal*/
   symbol7 v=none i=join line=21 w=1 c=black;

ods graphics on;
ods html;
proc gplot data=temp.usmap;
            title1 j=c h=3 c=black 'Funnel plot. 90-day delinquency rate in credit card loans in the USA';
            axis1 label=(c=black a=90 h=2.5 '90-day delinquency rate') value=(c=black)order=(-0.005 to 0.05 by 0.005)
/**/;
            axis2 label=(c=black h=2.5 'Loan volume')  value=(c=black);
            plot dlqrt*volume=1
                      dlqrt_95pct*volume=2
                      dlqrt_5pct*volume=3
                      dlqrt_90pct*volume=4
                      dlqrt_10pct*volume=5
                      dlqrt_us*volume=7
      /overlay
                       legend=legend1
      vaxis=axis1 haxis=axis2
             cframe = white
      hminor = 0
                       vminor =0
                       anno=temp.outliers;
            legend1 value=(justify=left) label=('Legend:' justify=left position=(middle left))
                               value=(h=2.5) frame;
   run;
   ods graphics off;


proc print data=temp.outliers noobs label;
            var STATECODE STATENAME volume dlqrt dlqrt_90pct;
            where flag_print ='Up-outlier';
            title 'Table. The US states with high delinquency rate';
run;
```

```
proc print data=temp.outliers noobs label;
            var STATECODE STATENAME volume dlqrt dlqrt_10pct;
            where flag_print='Down-outlier';
            title 'Table. The US states with low delinquency rate';
run;


/* Mapping the risk classification onto a chart*/
/* Constructing Chart 2 with risk classification exhibited on the USA map */
/* Auxiliary work: positioning states' name*/
/* This code is a part of Example 6: Example 6: Labeling the States on a U.S. Map
SAS-online documenation for proc gmap: http://support.sas.com/onlinedoc/913/docMainpage.jsp  */
data center;
  length function $ 8;
  retain flag 0 xsys ysys '2' hsys '3' when 'a' style 'swiss';
  set maps.uscenter (where=(fipstate(state) ne 'DC') drop=long lat);
  function='label';
  text=fipstate(state);
  size=2.5; position='5';
    if ocean='Y' then
    do;
                    position='6'; output;
      function='move'; flag=1;
    end;
    else if flag=1 then
    do;
      function='draw';
      size=.25; flag=0;
    end;
  output;
run;


/* Developing a risk map */
ods graphics on;
ods html;

proc gmap map=maps.us data=temp.usmap ;
            title1 j=c h=3 c=black 'Chart. Classification of credit card portfolios by 90-day delinquency rate in the USA';

            pattern1 value=solid color=red;        /* Extremely high*/
            pattern2 value=solid color=blue ;      /* Extremely low*/
            pattern3 value=solid color=pink;       /* Moderate high */
            pattern4 value=solid color=cyan;       /* Moderate low*/
            pattern5 value=solid color=lime;    /* Normal: green */
            legend1 value=(justify=left) label=('Risk' justify=left 'levels:' position=(middle left)) frame;
            id  STATECODE ;
            choro risk_level /levels=5 coutline=gray         annotate=center legend=legend1;
run;

ods graphics off;
```

References

1. David R. Anderson, Dennis J. Sweeney, Thomas A. Williams. Statistics for Business and Economics. 9th edition. Thomson South-Western, 2005
2. Dover Doug. Funnel Plots: Visualizing small area estimates. Presentation at SAS users group meeting in Edmonton, Canada: http://www.sas.com/offices/NA/canada/downloads/presentations/Edmonton2010/Funnel.pdf
3. PROC GMAP, Example 6: Labeling the States on a U.S. Map, SAS 9.1.3 on-line support documentation, http://support.sas.com/onlinedoc/913/docMainpage.jsp

Contact Information

Your comments and questions are valued and encouraged. Contact the author at:

Name: Yevhen Yankovskyy
Enterprise: ATB Financial
Address: 9888 Jasper Ave., Transit #426,
Edmonton, Alberta, Canada, T5J 1P1

Phone: (780) 392-9243
Fax: (780) 989-5301
E-mail: yyankovskyy@yahoo.com
Web: n/a