



Survival Analysis Workflow

Assessing the impact of Macro-Economic shocks on Credit Portfolios and predicting the Time of Default

Anupam Saha
SAS Research and Development (India) Pvt Ltd
Pune, India

Agenda

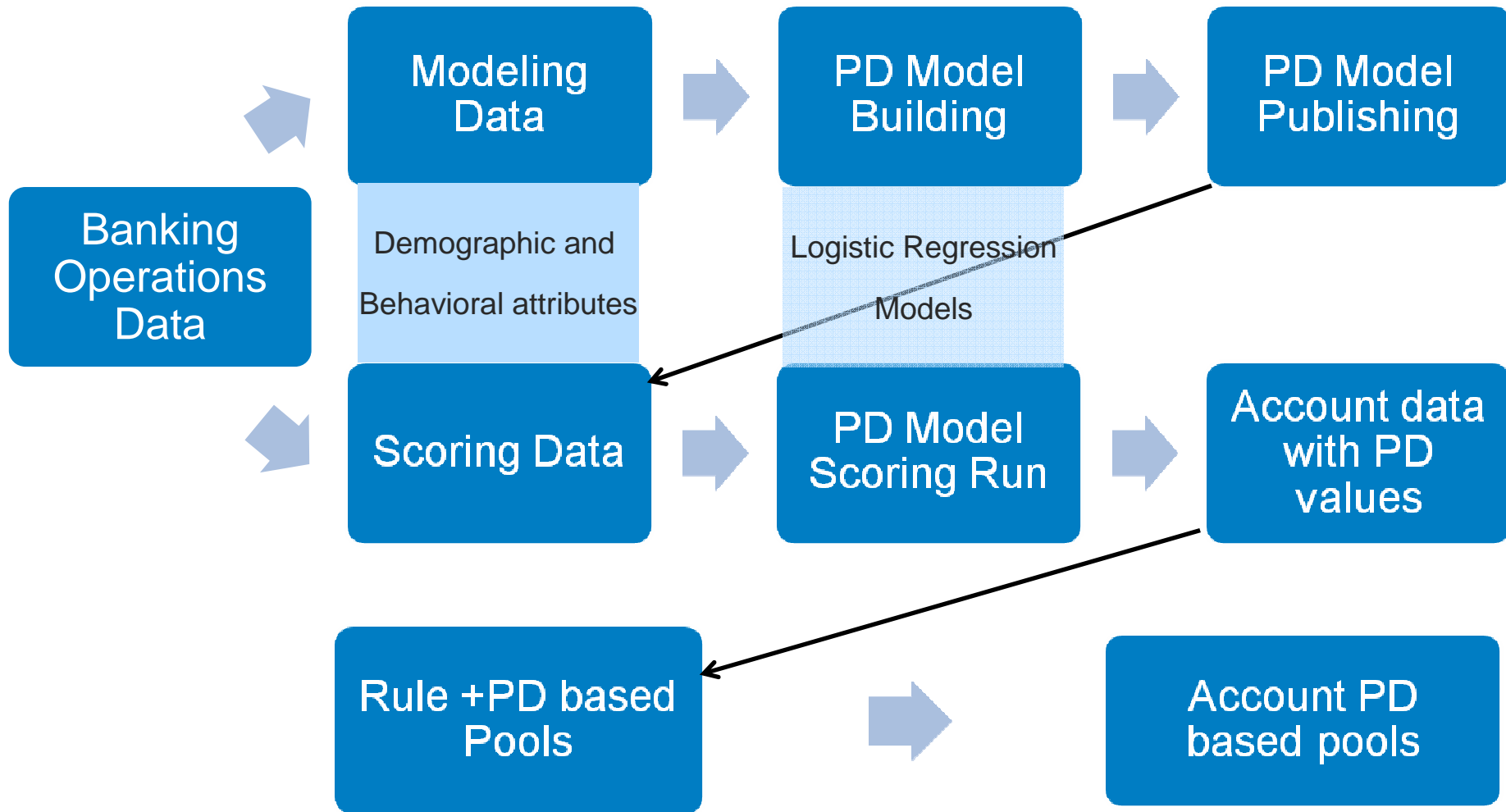
- Existing Methodology and Limitations
- Proposed Methodology
- Data Requirements
- Validation Techniques
- Benefits
- Conclusion and Challenges

Agenda

- Existing Methodology and Limitations
- Proposed Methodology
- Data Requirement
- Validation Techniques
- Benefits
- Conclusion and Challenges

Existing Methodology

Typical workflow of the Probability of Default estimation and segmentation of accounts based on PD values -



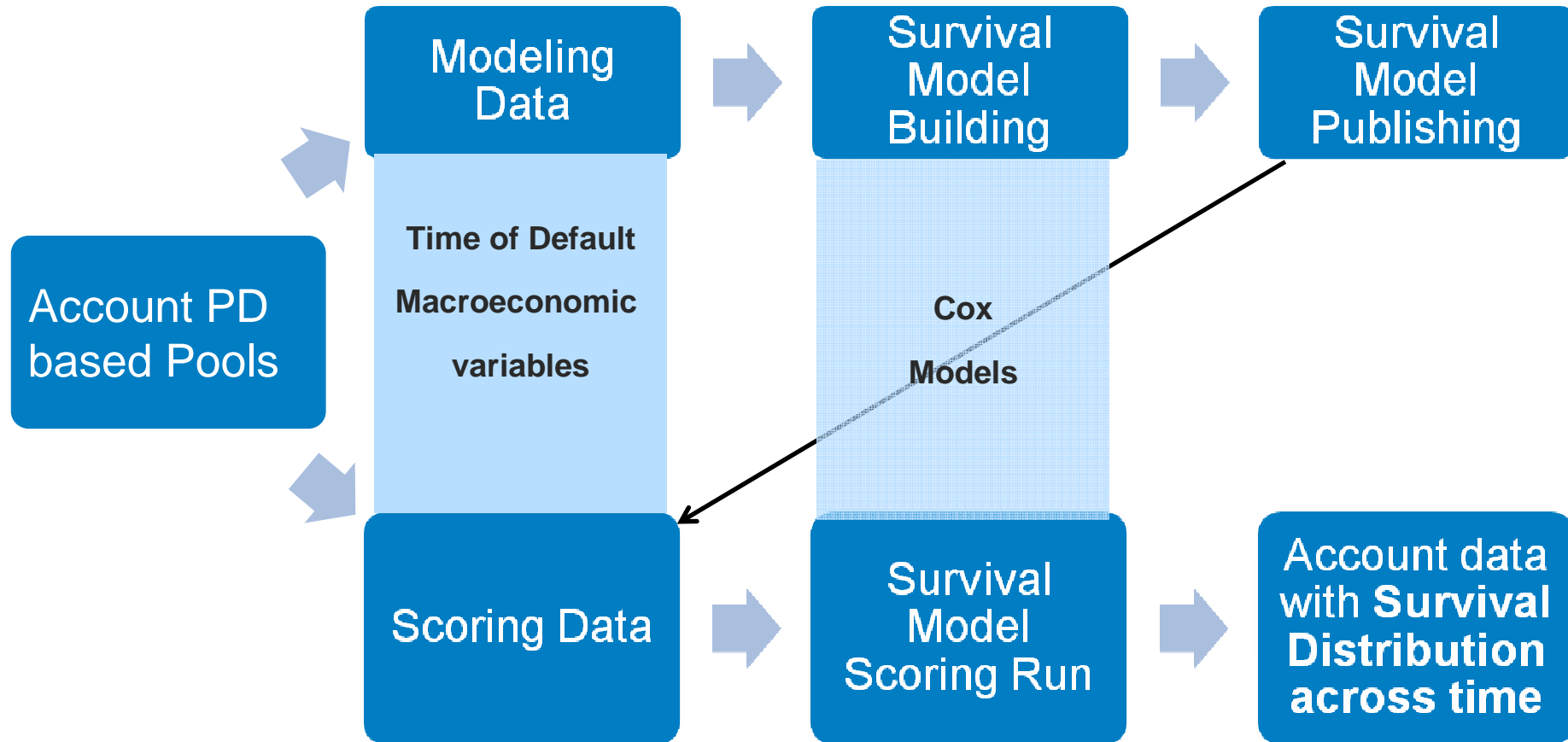
Limitations

- Impact of changes in Macro Economic variables is typically not considered in Logistic Regression based PD models
- Predicted Probability of Default is assumed to remain constant across Outcome Period
- Prediction of time to default is not calculated

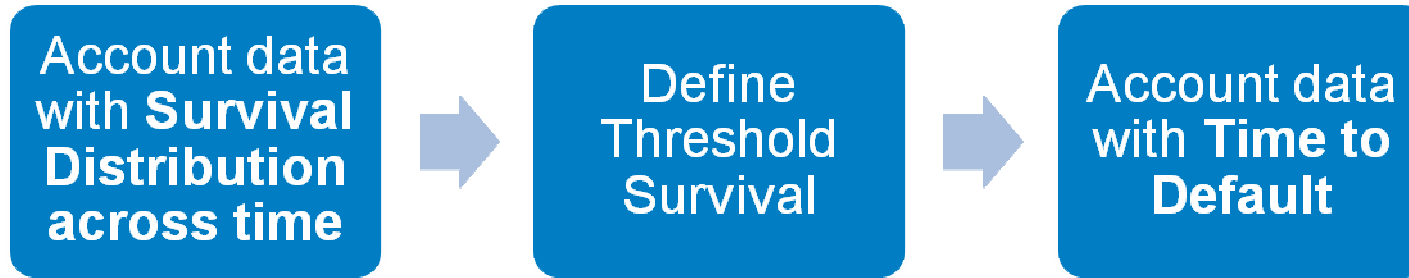
Agenda

- Existing Methodology and Limitations
- Proposed Methodology
- Data Requirements
- Validation Techniques
- Benefits
- Conclusion and Challenges

Proposed Methodology



Proposed Methodology



- Additional derivations –
 - Survival distribution of default at pool level across time
 - Default count distribution across pools across time

Cox Model

The Cox model formula states:

$$h_i(t) = h_0(t) \exp(\beta_1 X_{i1}(t) + \beta_2 X_{i2}(t) + \dots + \beta_k X_{ik}(t))$$

- Where,

i is a subscript for observation, $i=1(1)n$

$h_i(t)$ = Hazard of the i th individual at time t (*account*)

$h_0(t)$ = Baseline hazard function

$X_{i1}(t) \dots X_{ik}(t)$ are the k time-dependent covariates of the i th individual at time t (*Macro economic variables*)

$\beta_1 \dots \beta_k$ are the regression parameters associated with the covariates, estimated by the Partial Likelihood method in Cox model

Implementation of Cox Model

- SAS Phreg procedure performs Cox model

Account id	Time	Default	Cov1(Covariate at time 1)	Cov1(covariate at time 2)	Cov1(covariate at time 3)	Cov1(covariate at time 4)
1	2	1	<value>	<value>		
2	3	1	<value>	<value>	<value>	
3	4	0	<value>	<value>	<value>	<value>

Modeling Data

Account id	Time	Default	T1	T2	Status	Covariate
2	3	1	0	1	0	<value>
2	3	1	1	2	0	<value>
2	3	1	2	3	1	<value>

Modeling Data transformed to Counting process style of input

Perform Proc Phreg

Account id	Time	Survival Probability
1	1	<value>
1	2	<value>
1	3	<value>
1	4	<value>
1	5	<value>

Account level Survival distribution

- For counting process style of input data, the covariate values remain unchanged in a particular time interval

Piece Wise Constant Baseline Hazard model

- An alternative to Cox model implementation is Piece Wise Constant Baseline Hazard model. SAS Phreg procedure supports this model as well.
- By piece wise hazard model

$$h_i(t) = h_0(t) \exp(\beta'x_i)$$

Where, $h_0(t) = \lambda_j$ $a_{j-1} \leq t < a_j$ ($j = 1, \dots, J$)

- The baseline cumulative hazard function is denoted by

$$H_0(t) = \sum_{j=1}^J \lambda_j \Delta_j(t)$$

Where,
$$\Delta_j(t) = \begin{cases} 0 & t < a_{j-1} \\ t - a_{j-1} & a_{j-1} \leq t < a_j \\ a_j - a_{j-1} & t \geq a_j \end{cases}$$

Agenda

- Existing Methodology and Limitations
- Proposed Methodology
- Data Requirements
- Validation Techniques
- Benefits
- Conclusion and Challenges

Data Requirements

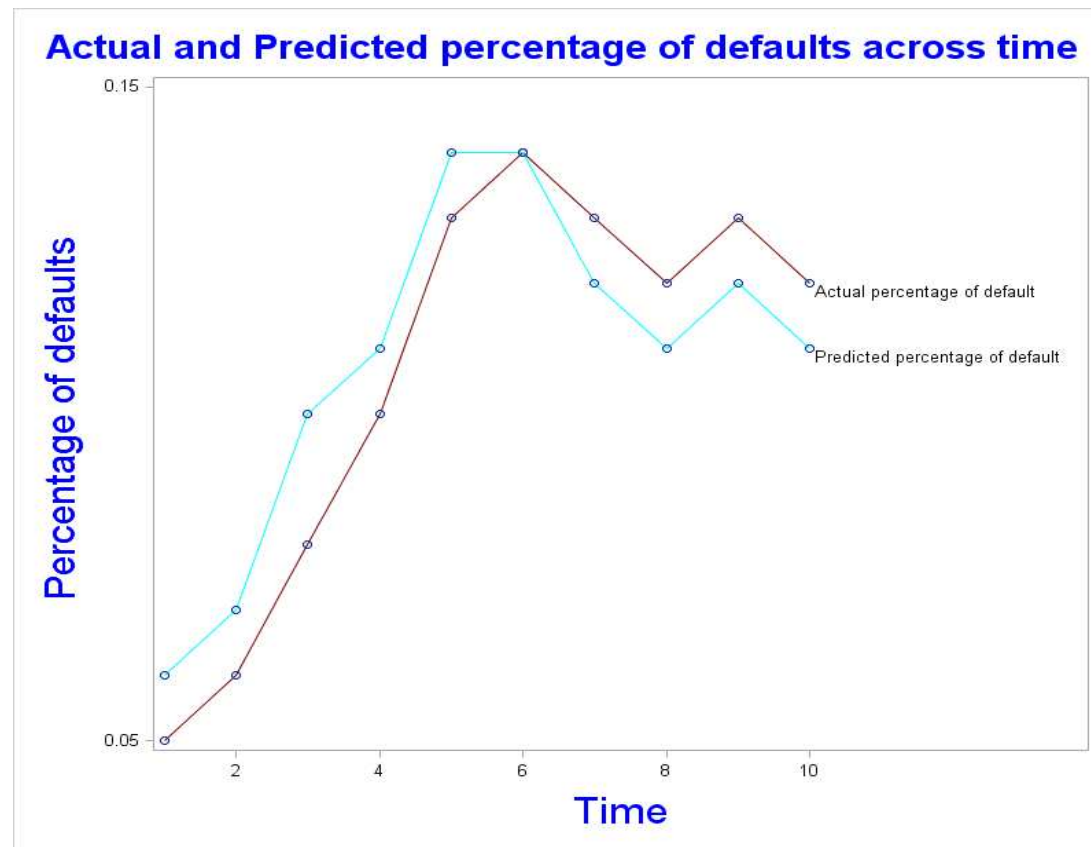
- Default Time of Each Defaulted Customer
- Demographic Variables (income, gender, location, etc)
- Behavioral variables (ie, account performance related variables)
- Macro Economic Variables (at lowest granularity possible)
 - Inflation Rate
 - Unemployment Rate/Employment Growth Rate
 - Housing Price Index
 - Consumer Price Index
 - Average Personal Income Growth Rate / Growth Rate
- Macro Economic Variables captured at
 - Account default date
 - Prior to default date
 - Loan start date
 - Regular time intervals during Observation Period

Agenda

- Existing Methodology and Limitations
- Proposed Methodology
- Data Requirements
- Validation Techniques
- Benefits
- Conclusion and Challenges

Validation Techniques

- Compare development and validation dataset
- Compare actual and predicted percentage of defaults across time



Validation Techniques

- Calculation of MSE:

$$MSE = \frac{\sum_{i=1}^n (\text{Actual time of default} - \text{predicted time of default})^2}{n}$$

Where, n is the total number of accounts

- Implementation of Brier score
- Implementation of Non-parametric median test

Agenda

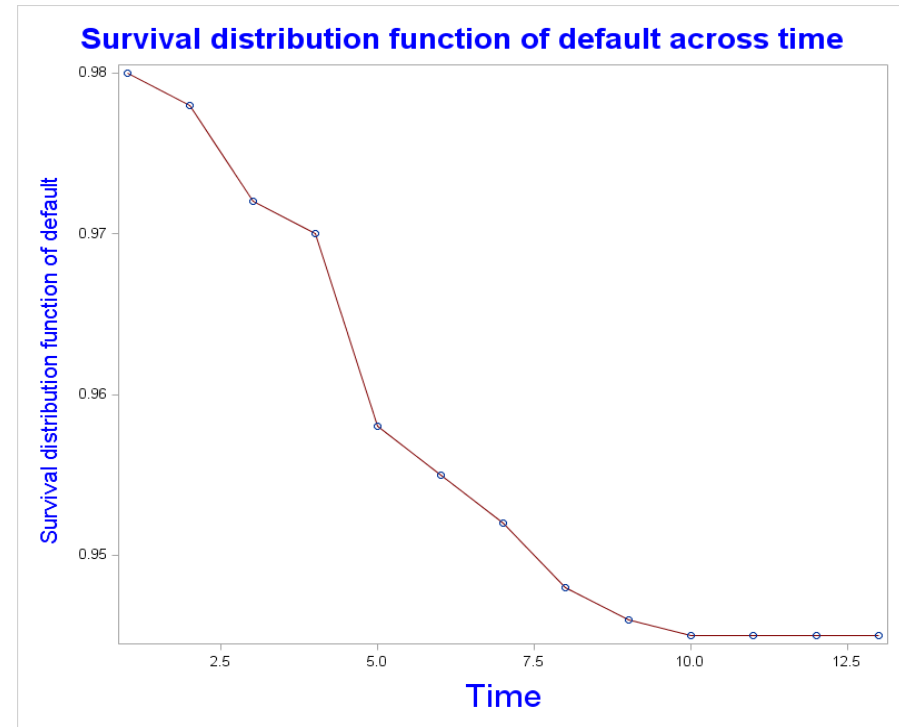
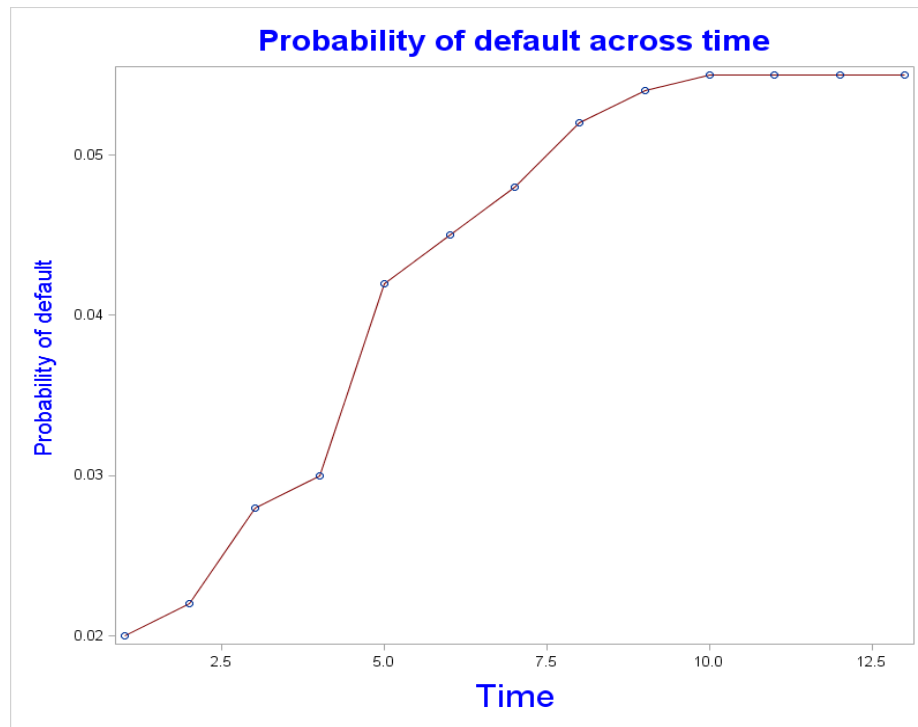
- Existing Methodology and Limitations
- Proposed Methodology
- Data Requirements
- Validation Techniques
- **Benefits**
- Conclusion and Challenges

Benefits

- Advantage of including time-varying covariates
- Capability to capture the change in default behavior due changes into macro economic values
- Capability to determine the expected time of default for an account

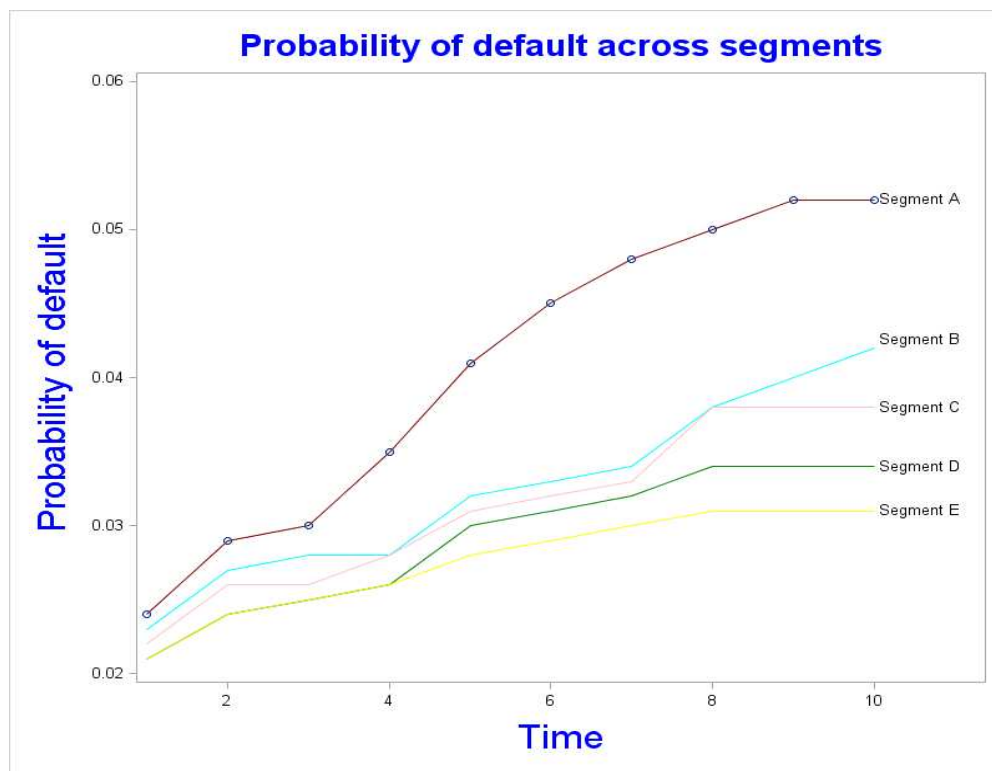
Benefits

- Predicting probability of default across time in future
- Predicting Survival probability of default for future time points



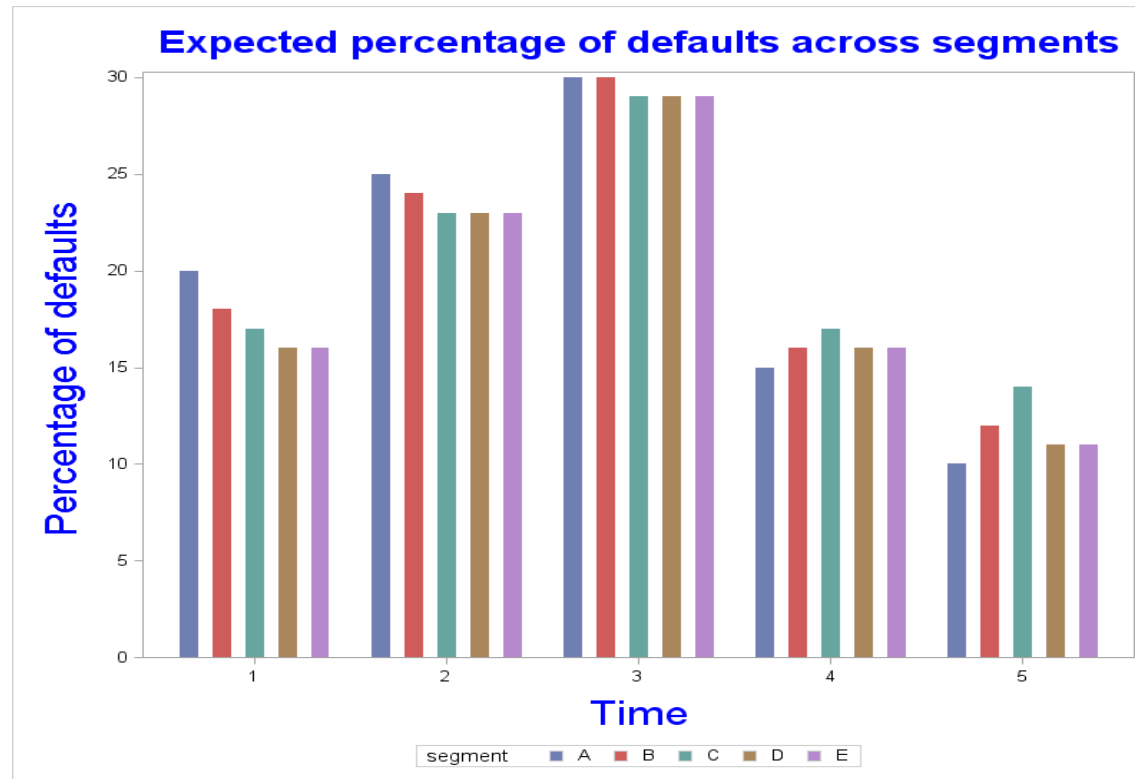
Benefits

- Probability of default across segments

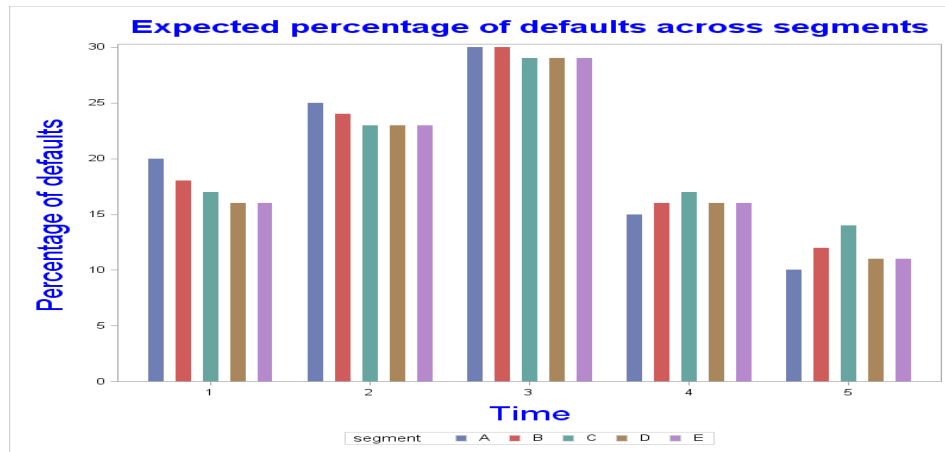


Benefits

- Prediction of expected percentage of defaults in future time periods for each segments



Benefits : Scenario Analysis

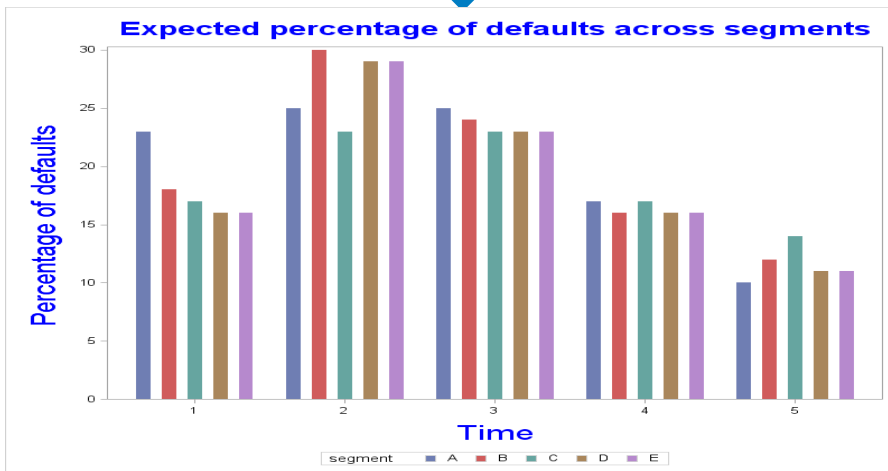


Changed covariate values

<value>
<value>
<value>
<value>
<value>

Capture expected default time

Recalculate Survival function of default



Change in default distributions across segments

Agenda

- Existing Methodology and Limitations
- Proposed Methodology
- Data Requirements
- Validation Techniques
- Benefits
- Conclusion and Challenges

Conclusion and Challenges

- Conclusion
 - Using Cox model, it is possible to study the impact of macro economic variables on probability of default
 - Generating survival probabilities across time
 - Predicting the time to default
- Challenges
 - Capturing Macro Economic variables at granular level
 - Correlation/collienarity
 - Different time lag in impact of different macro variables e.g. interest rate change vs manufacturing output change.



Thank you!

NS12

Slide 25

NS12

Add a slide before this to discuss any issues/drawbacks/gaps etc. to make people see that you have considered all angles. what are the problems with this approach ?

Naeem Siddiqi, 05/08/2011

Multiplicative Hazard Model

- Consider a set of subjects such that the counting process $N_i \equiv \{N_i(t), t \geq 0\}$ for the i th subject represents the number of observed events experienced over time. The sample paths of the process are step functions with jumps of size +1, with, $N_i(0) = 0$

Let β denote the vector of unknown regression coefficients. The multiplicative hazards function for $\Lambda_i(t, \mathbf{Z}_i(t))$ is given by

$$Y_i(t) d\Lambda_i(t, \mathbf{Z}_i(t)) = Y_i(t) \exp(\beta' \mathbf{Z}_i(t)) d\Lambda_0(t)$$

where, $Y_i(t)$ indicates whether the i th account defaults at time t (specifically, $Y_i(t) = 1$ if defaults, otherwise $Y_i(t) = 0$)

$\mathbf{Z}_i(t)$ is the vector of explanatory variables for the i th account at time t .

$\Lambda_0(t)$ is an unspecified baseline hazard function.

Overview of Survival Analysis Techniques

- The survival analysis technique provides hazard function

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr (t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

- The survival analysis technique provides survival function

$$S(t) = \exp\left\{-\int_0^t h(u) du\right\}$$

Resources

- Paul D. Allison. Survival Analysis using SAS A Practical Guide
- SAS/STAT 9.2 User's Guide
- A Paper "Credit Scoring With Macro Economic Variables Using Survival Analysis." by Tony Bellotti and Jonathan Crook
- Fleming, T. R., and Harrington, D. P. (1991), Counting Processes and Survival Analysis, New York: Wiley.
- Anderson, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993), Statistical Models Based on Counting Processes, New York: Springer-Verlag.
- For Generating survival function refer to (Kalbfleisch and Prentice; [1980](#), pp. 84–86)
- For Multiplicative Hazard Model refer to Fleming and Harrington ([1991](#)) and Andersen et al. ([1992](#)).