Incorporating Stochastic Dominance Constraints in Credit Scoring Methodology – Models & Experimental Results.

Debashish Sarkar
Management Sciences Group
CIT, New Jersey, USA

Abstract

Stochastic dominance is a desirable property for all credit scoring models. Typically credit-scoring formulations model the response variable (good or bad) as a logistic function of the predictor variables and the parameter estimates are obtained by maximizing the log likelihood function (alternatively, some distance function between the good and bad scores is optimized.) The outputs of such methodology include the cumulative distribution functions (CDFs) of the good and bad population scores, and the effectiveness of the credit scoring function is measured by the "separation" between these CDFs. Stochastic dominance is one such measure of separation. The scoring function is said to have first order stochastic dominance (FSD) property if the CDF for the bad accounts lies above the good CDF curve for all scores. Similarly, the scoring function satisfies second order stochastic dominance (SSD) property if, for any given score value, the area to the left under the bad CDF curve is greater than or equal to the value for the good CDF curve. Credit scoring functions, in practice, exhibit FSD property (which is stronger than SSD) almost always even though they are not explicitly modeled for. For some applications, FSD or SSD property does not hold for a very small range of scores (typically at the low end). Direct optimization of metrics that involve point-wise comparison of the good and bad CDF curves has been challenging because of the computational complexity involved in modeling the CDFs as a function of individual scores within the optimization setup. In this paper, using Cornish-Fisher quantile approximation and a duality result on SSD, we show how to explicitly incorporate FSD or SSD constraints in the credit scoring model development process.

We use Cornish-Fisher approximation to represent the quantile-distribution of the good and the bad scores (that is, the inverse of the good and bad CDFs.) The Cornish-Fisher quantile approximation formula involves expressions for skewness & excess kurtosis of the good and bad scores. FSD constraints can be enforced in a straightforward way in terms of quantile functions. SSD constrains are modeled in terms of area under quantile functions following a duality result from Ogryczak and Ruszczyski. Implementation details and results from several practical problems will be shown.