Marginal Information Value

Andrija Djurovic

www.linkedin.com/in/andrija-djurovic

Probability of Default Modeling and Stepwise Procedures

- The stepwise logistic regression procedures most commonly used for modeling the PD still rely on the p-values of the coefficients.

- An alternative, less frequently employed in practice, is based on Marginal Information Values (MIV).

- Can we benefit from combining them?

Marginal Information Value (MIV)

- The concept behind the MIV score for a new variable is that, after constructing a model using a specific set of available variables, the MIV evaluates the new variable by measuring the additional information it is likely to contribute compared to the predictions generated by the current model.

- Include a sufficient number of variables to account for the variation in outcomes across the sample, but avoid adding unnecessary ones.

Example of MIV Calculation

MIV table for the variable Maturity

Maturity	Observed			Expected		
	# good obs.	# bad obs.	WoE obs.	# good exp.	# bad exp.	WoE exp.
[4,7]	78	9	1.31	62.29	24.71	0.08
(7,15]	264	80	0.35	240.20	103.80	-0.01
(15,42]	328	171	-0.20	348.60	150.40	-0.01
(42,Inf]	30	40	-1.13	48.91	21.09	-0.01

MIV score

```
delta = "WoE obs." - "WoE exp."
miv_g = sum("# good obs." * delta) / sum("# good obs.")
miv_b = sum("# bad obs." * delta) / sum("# bad obs.")
miv = miv_g - miv_b
miv
## [1] 0.2611226
```

MIV Threshold and Supporting Statistical Test

- The most frequently used MIV threshold is 0.02.

- Sometimes the absolute MIV threshold is supported by the Marginal Chi-square test.