Understanding Variable Interactions for Better Scorecards

Presented By Sam Buttrey,
Portfolio Defense Consulting Group
2005 Annual Conference

PORTFOLIO DEFENSE

Some Notes on Interactions

- An "interaction" is when the effect of one variable on a response depends on the level of a second variable

- Example: effect on health of (alcohol) + effect of (drug) ≠ effect of (alcohol + drug)

- Example: effect on creditworthiness associated with one 90-day delinquency is larger when income is high (let's say)

Interactions Are Often Ignored

- In scorecard models, interactions are often ignored. The number of points awarded for one level of one variable is constant, regardless of the values of any other variables

- Experience suggests that a small number of well-chosen interactions can improve model fit substantially

PORTFOLIO DEFENSE

Modeling Interaction

- Logistic regression (widely used, interpretable; interactions must be specified explicitly)

- Classification trees (interactions handled automatically; less statistical theory)

- Neural networks (interactions handled implicitly; no interpretation)

PORTFOLIO DEFENSE

Scorecard With No Interaction

- Household Income   Delinquent TLs

  $\leq$ 29,999   :   0 Pts   0: +55 Pts

  30k-49,999: 20 Pts   1: +45 Pts

  50k-79,999: 35 Pts   2-3: + 5 Pts

  $\geq$ 80,000   : 70 Pts $\geq$ 4:   0 Pts

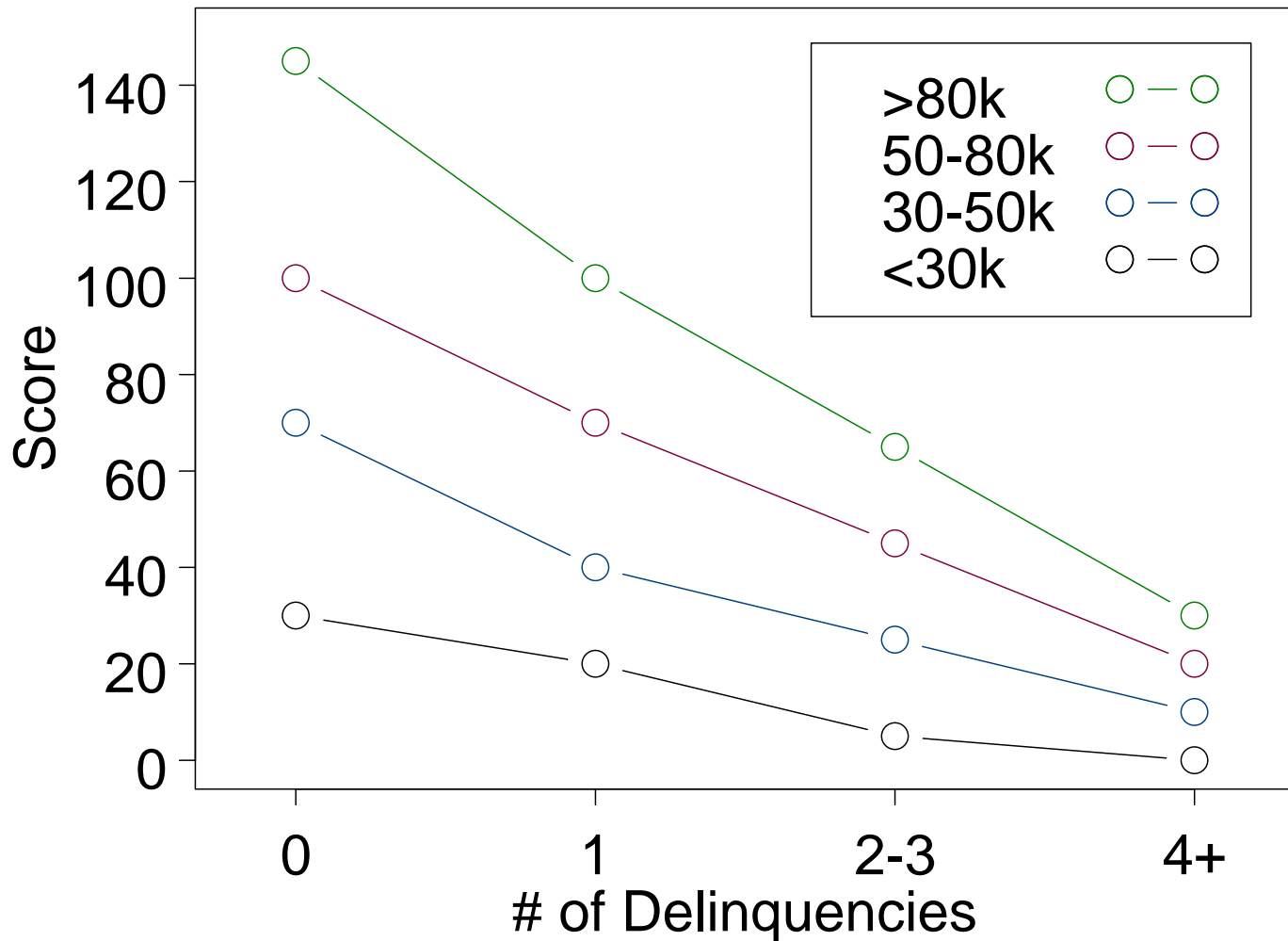- Going from the "0 delq" group to the "1 delq" group costs you 10 points, regardless of income

Scorecard With No Interaction
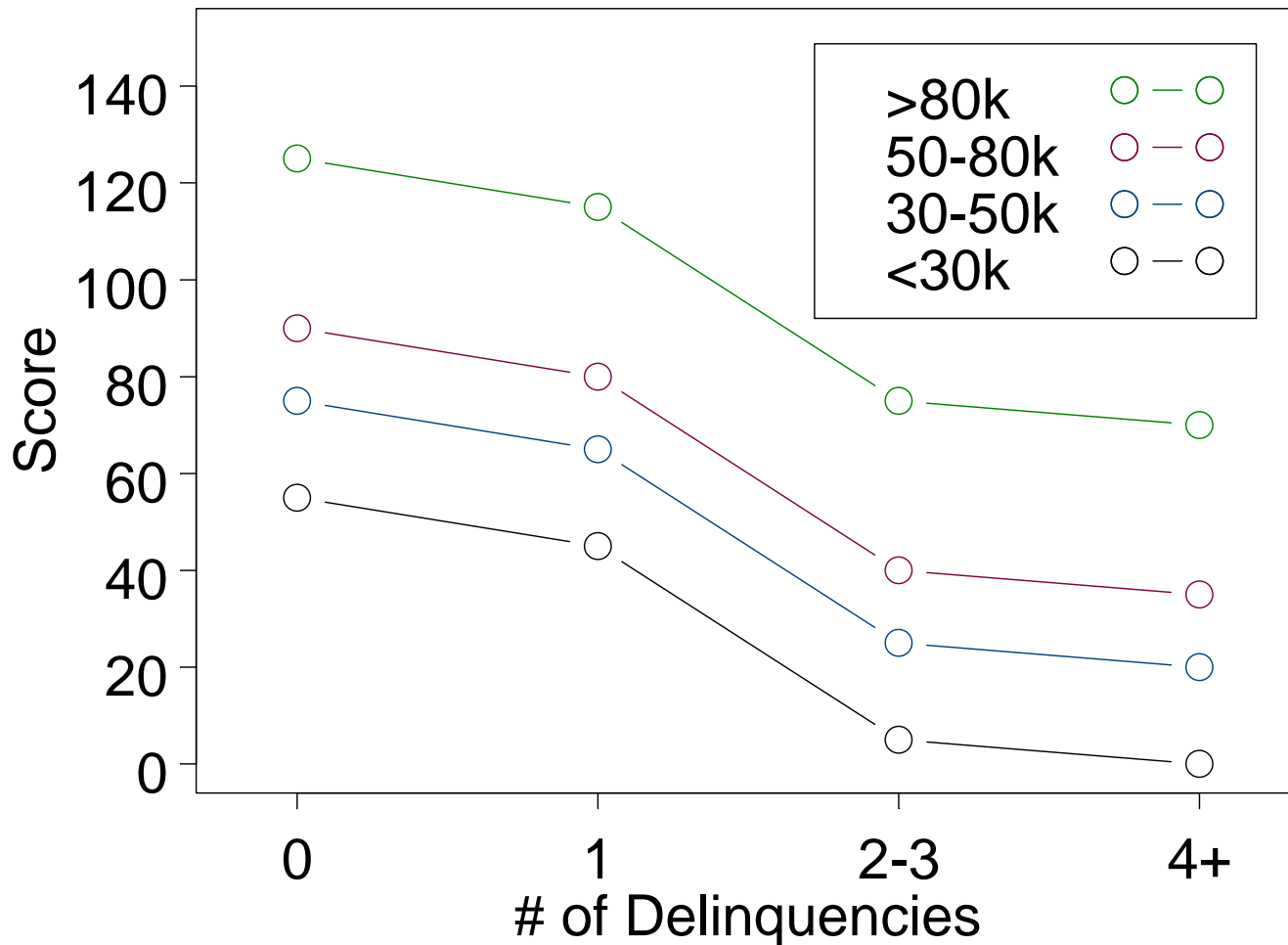
Scorecard With Interaction

Points	0 delq	1 delq	2-3 delq	$\geq$ 4 delq
< 30k	30	20	5	0
30-50 k	70	40	25	10
50-80k	100	70	45	20
$\geq$ 80k	145	100	65	30

PORTFOLIO DEFENSE

Scorecard With Interaction

Where Can This Be Used?

- Anywhere interaction exists and isn't currently being modeled
- Potentially large gains in predictive power, in exchange for some loss in interpretability
- Logistic regression techniques and theory continue to apply; other models also available

PORTFOLIO DEFENSE

Segmentation

- One obvious use: segmenting populations into scorecard segments

- We might expect the effect of some variables to be different across segments

- That is, we expect an interaction between that variable and segment

PORTFOLIO DEFENSE

Classification Trees

- This is a statistical model that identifies interactions

- Particularly useful for segmenting population into smaller, more homogeneous groups

- Start with observations in one "box"

- Split into two pieces so as to maximize homogeneity of outcome

Automatic Interaction Detection

- The "Trades < 4" group is divided by "EverDQ" (Yes or No), but the "Trades <= 4" group is divided by "#30+ < 2" (Yes or No)
- That is, the effect of "EverDQ" is bigger than "#30 < 2" for "Trades < 4," but smaller for "Trades >= 4"
- This is an interaction!

PORTFOLIO
DEFENSE

Conclusion

- Interactions exist in real data, so we should include them in the model

- They show how the effect of one variable on the response can depend on the value of another

- Useful for segmentation, potentially for scorecard construction