Midwestern Actuarial Forum

Finding Interactions

Serhat Guven, FCAS, MAAA

March 28, 2011

TOWERS WATSON

Agenda

- Background
- Balance
- Data Mining
- Saddles
- Conclusion

Background

- Simple model: relationship between rating levels of one factor is constant for all levels of other rating variables
- Assume two rating variables
  - Age: Youthful, Adult (Base), Mature, Senior
  - Gender: Male (Base), Female

Simple Model: Age + Gender

	Male	Female
Youthful	$\beta_0 + \beta_Y$	$\beta_0 + \beta_Y + \beta_F$
Adult	β_0	$\beta_0 + \beta_F$
Mature	$\beta_0 + \beta_M$	$\beta_0 + \beta_M + \beta_F$
Seniors	$\beta_0 + \beta_S$	$\beta_0 + \beta_S + \beta_F$

5 Parameters

Background

- Complex model: relationship between rating levels of one rating factor is different for all levels of another rating variable

Simple Model: Age + Gender

	Male	Female
Youthful	$\beta_0 + \beta_Y$	$\beta_0 + \beta_Y + \beta_F$
Adult	β_0	$\beta_0 + \beta_F$
Mature	$\beta_0 + \beta_M$	$\beta_0 + \beta_M + \beta_F$
Seniors	$\beta_0 + \beta_S$	$\beta_0 + \beta_S + \beta_F$

5 Parameters

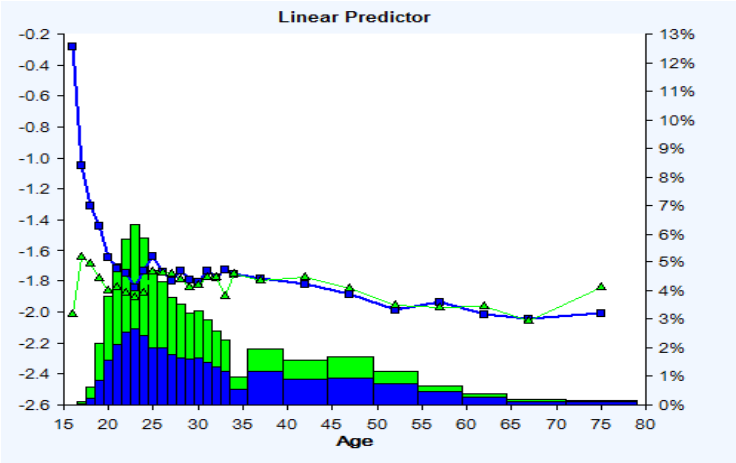
Interaction: Age + Gender+Age.Gender

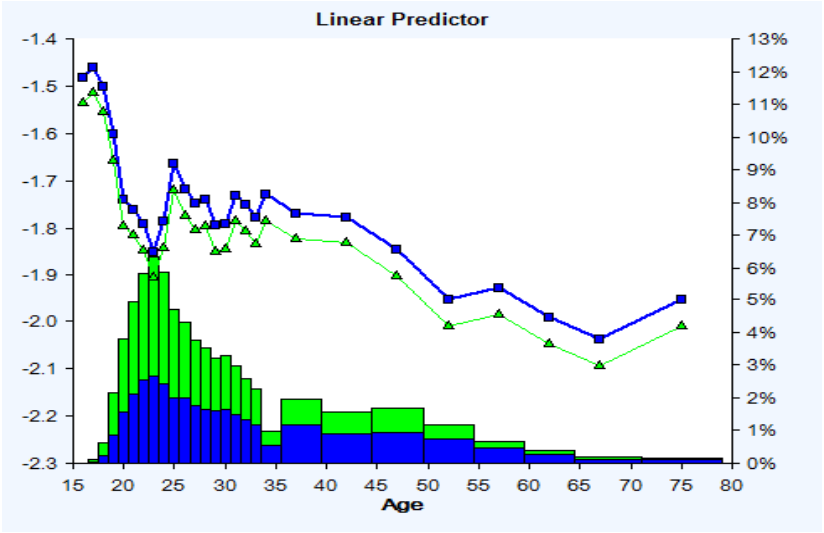
	Male	Female
Youthful	$\beta_0 + \beta_Y$	$\beta_0 + \beta_Y + \beta_F + \beta_{YF}$
Adult	β_0	$\beta_0 + \beta_F$
Mature	$\beta_0 + \beta_M$	$\beta_0 + \beta_M + \beta_F + \beta_{MF}$
Seniors	$\beta_0 + \beta_S$	$\beta_0 + \beta_S + \beta_F + \beta_{SF}$

8 Parameters

Background

- “Parallel” lines from the simple model





- Interaction breaks the parallel assumption

Background

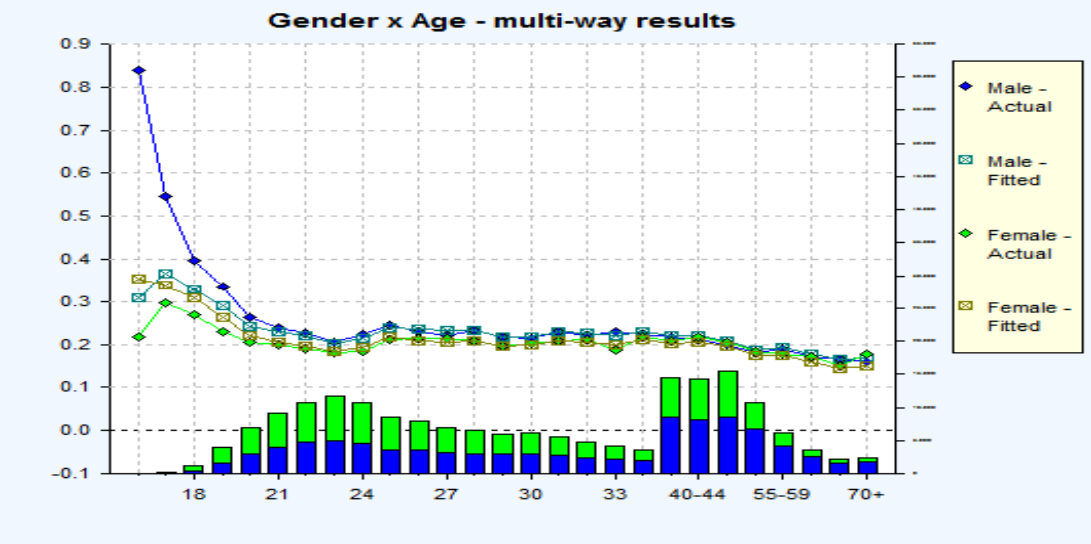
- Interaction challenges
  - Detection
  - Simplification
  - Verification

- Focus for today is the detection of the interaction construct

- Need a systematic approach to search the space for interactions
  - Balance
  - Data Mining
  - Saddles

Balance

- Approach is to compare aggregate average observed values vs. aggregate average fitted values



- Imbalance suggests need for interaction

Balance

- Given the following
  - Observed value

$$y_i = \frac{\text{Claims}_i}{\text{Exposures}_i}$$

  - Fitted value (assumes simple model structure)

$$\hat{y}_i = \mu_i = h(x_i\beta)$$

  - Weighted average observed value and weighted average fitted values for Class k

$$A_k = \frac{\sum_{i \in k} y_i \text{ x Exposures}_i}{\sum_{i \in k} \text{Exposures}_i} \qquad \hat{E}_k = \frac{\sum_{i \in k} \hat{y}_i \text{ x Exposures}_i}{\sum_{i \in k} \text{Exposures}_i}$$

Balance

- For each combination of cells for two rating factors derive the following:

	Male	Female
Youthful	D_{YM}	D_{YF}
Adult	D_{AM}	D_{AF}
Mature	D_{MM}	D_{MF}
Seniors	D_{SM}	D_{SF}

- Such that:

$$D_k = \frac{\text{Exposures}_k \text{ x } (A_k - E_k)^2}{E_k}$$

- Then:

$$Q = \sum_{\text{Age}} \sum_{\text{Gender}} D_k$$

Follows a chi squared distribution with (n-1)*(m-1) degrees of freedom

Balance

- Chi squared test then run for every two way combination
- Framework allows for ranking of potential constructs

Rank	Factor 1	Factor 2	Chi Test
1	Driving Restriction	Age	0.0000
2	Age	Gender	0.0000
3	Driving Restriction	NCD	0.0000
4	NCD	Gender	0.0000
5	NCD	Age	0.0001
6	Protected NCD	Gender	0.0002
7	Driving Restriction	Gender	0.0004
8	Driving Restriction	Protected NCD	0.0006
9	LossYear	Driving Restriction	0.0008
10	LossYear	Gender	0.0132
11	Vehicle Age	NCD	0.0176
12	Driving Restriction	Vehicle Category	0.0195
13	LossYear	Protected NCD	0.0425
14	Vehicle Category	Gender	0.0670
15	Rating Area	Gender	0.1185
…	…	…	…

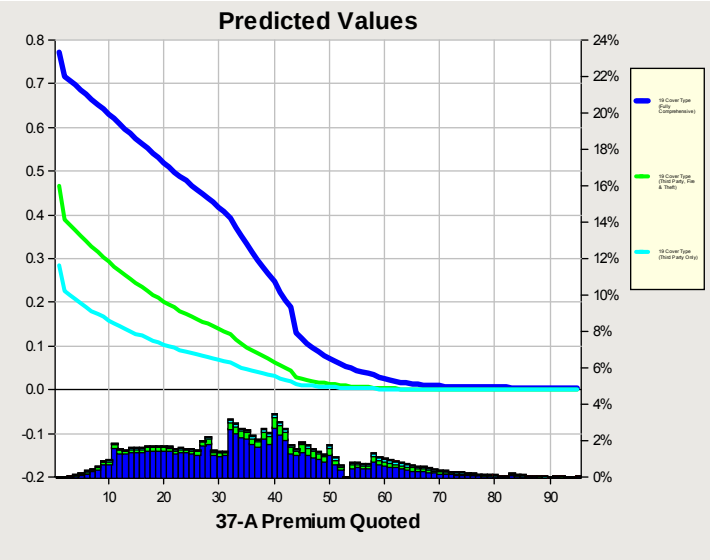
Balance

- Advantages
  - Can quickly identify areas in the model where interactions are needed
  - “Exponential” effect of distributional biases can magnify the importance of one interaction structure vs. another

- Disadvantages
  - Sensitive to noise of severity
  - Limited guidance as to simplification

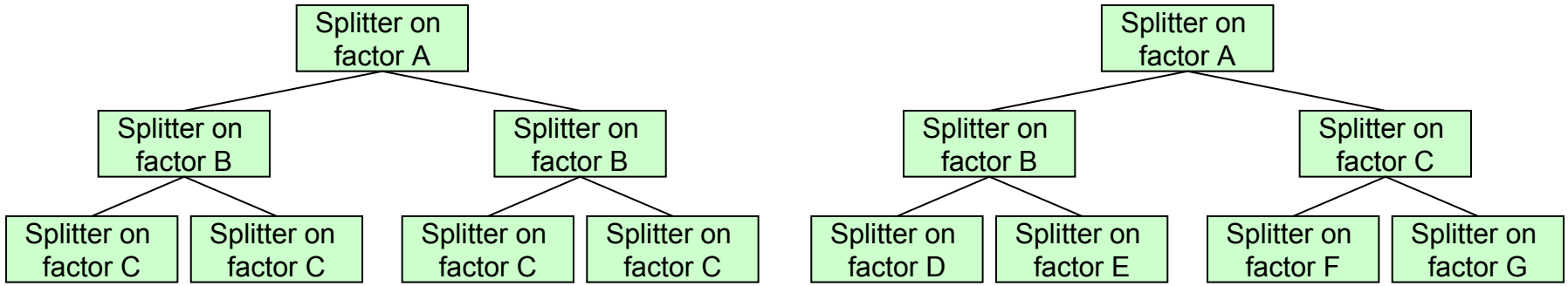
Data Mining

- CART can be used to help identify interactions.
- The diagram shows a new business conversion model.
- The shape of the conversion curves differs drastically for Comp Only policies compared to Liability and Full Coverage policies



Data Mining

- What would an interaction look like in CART?



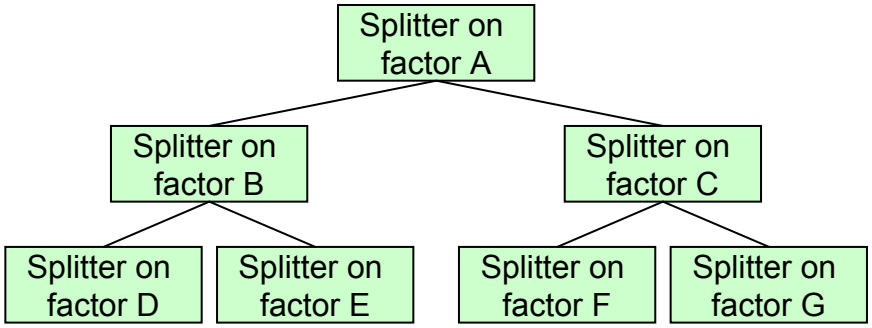
- The left hand tree has similar structures down each branch and so is unlikely to indicate interactions
- The right hand tree has different structures depending on which branch is traversed. This might have interactions.

Data Mining

- CART will not guarantee interactions

- CART can provides additional clues as to what interactions to test.
- This tree may have the following interactions

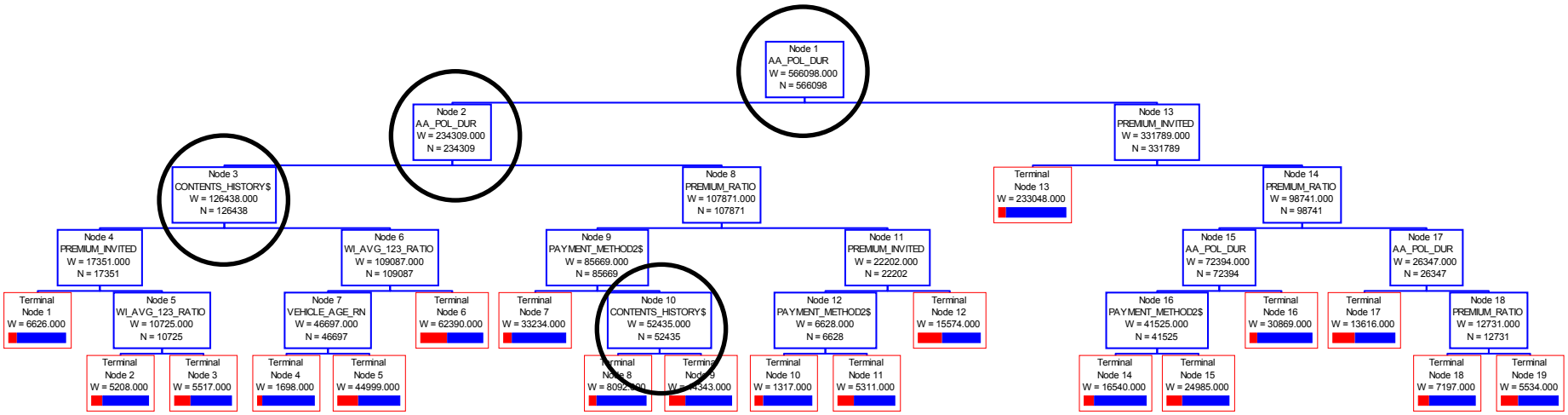
  - A x B, A x C, A x D, A x E, A x F,
  - A x G, B x D, B x E, C x F, C x G,
  - A x B x D, A x B x E,
  - A x C x F, A x C x G



- The list of candidate interactions can grow quickly!

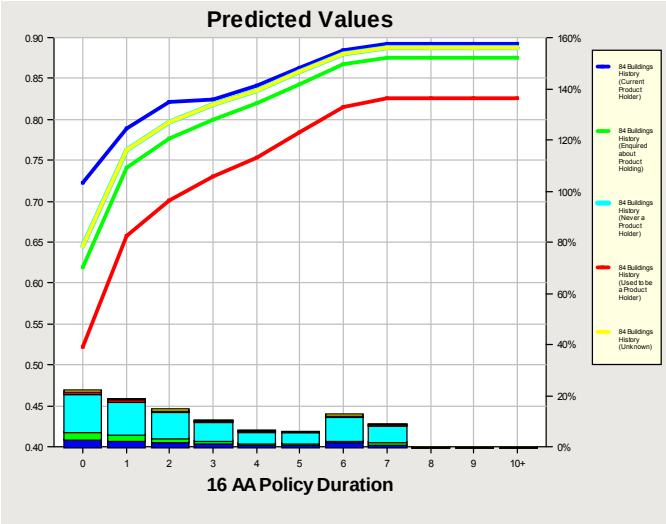
Data Mining

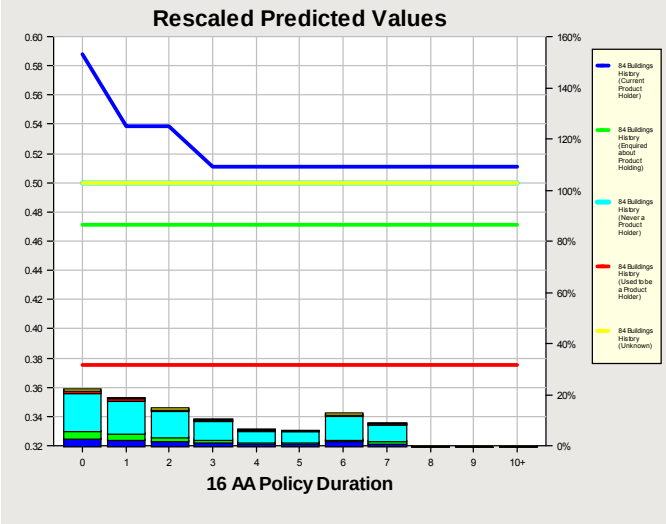
- Case study: auto renewals model
  - Balance test identified 23 interactions
  - Tree identified cross holdings and tenure interaction not from the balance test



Data Mining

- Additional analysis validated the interaction identified by CART
- Policyholders with multiple lines more likely to renew for the low duration policies



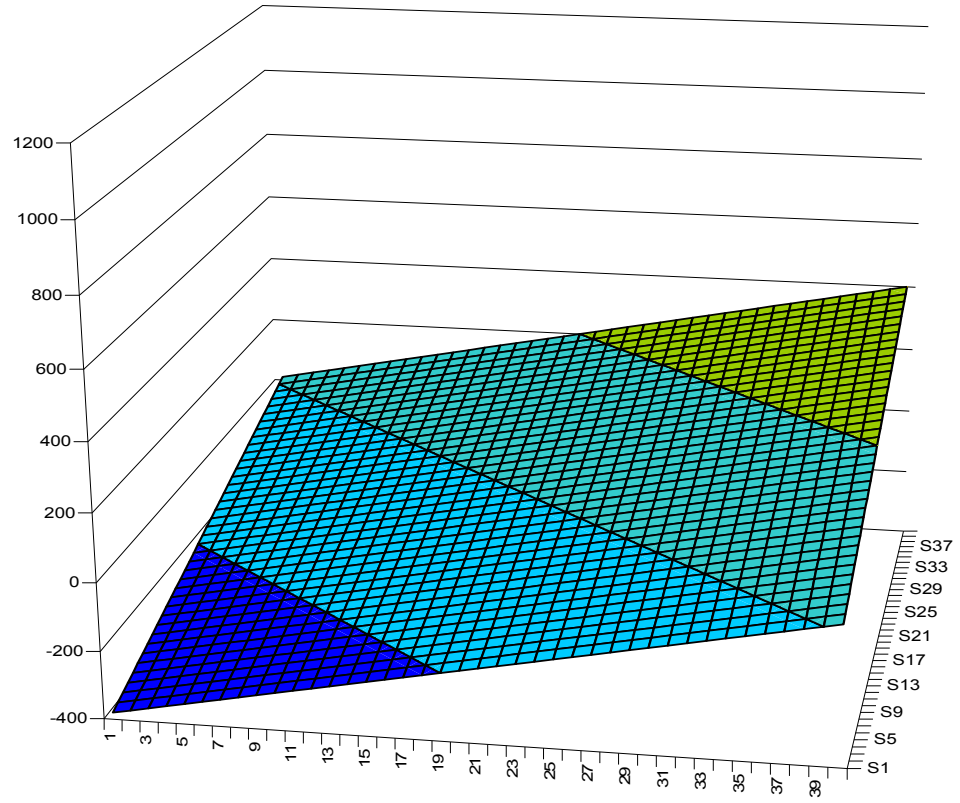


Data Mining

- Advantages
  - Quickly identify potential n-way interactions
  - Suggests areas of localization
- Disadvantages
  - Growth in complexity
  - Better performance when response is structure as a discrete (i.e. binomial/multinomial) construct
  - Limited guidance as to simplification

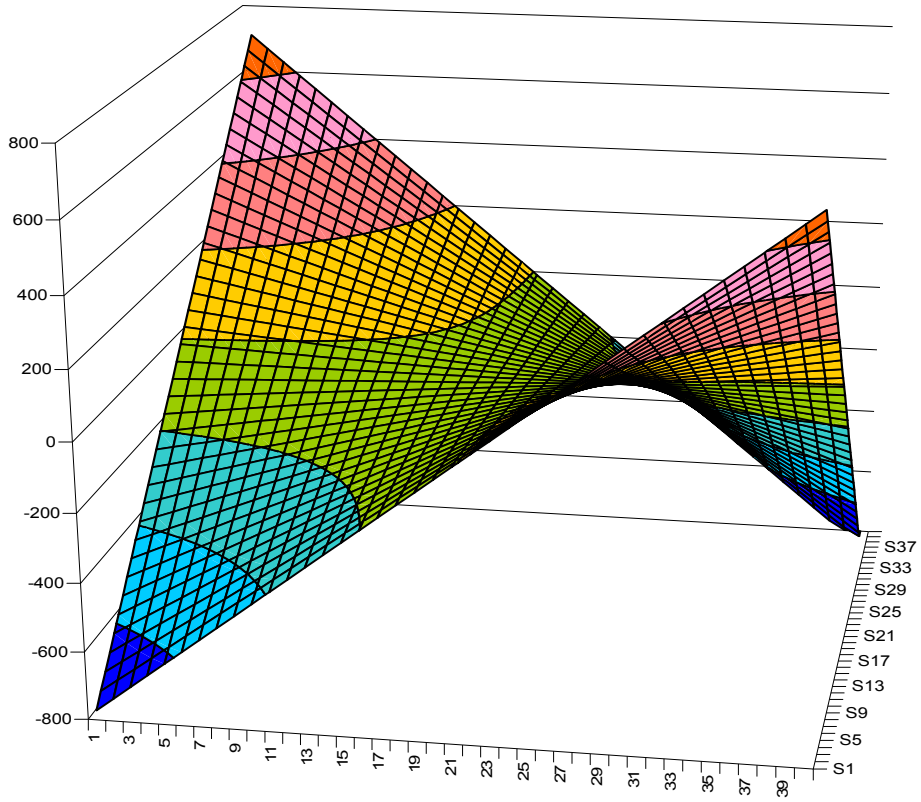
Saddles

- Quadrant saddle: revisiting an simple main effect model



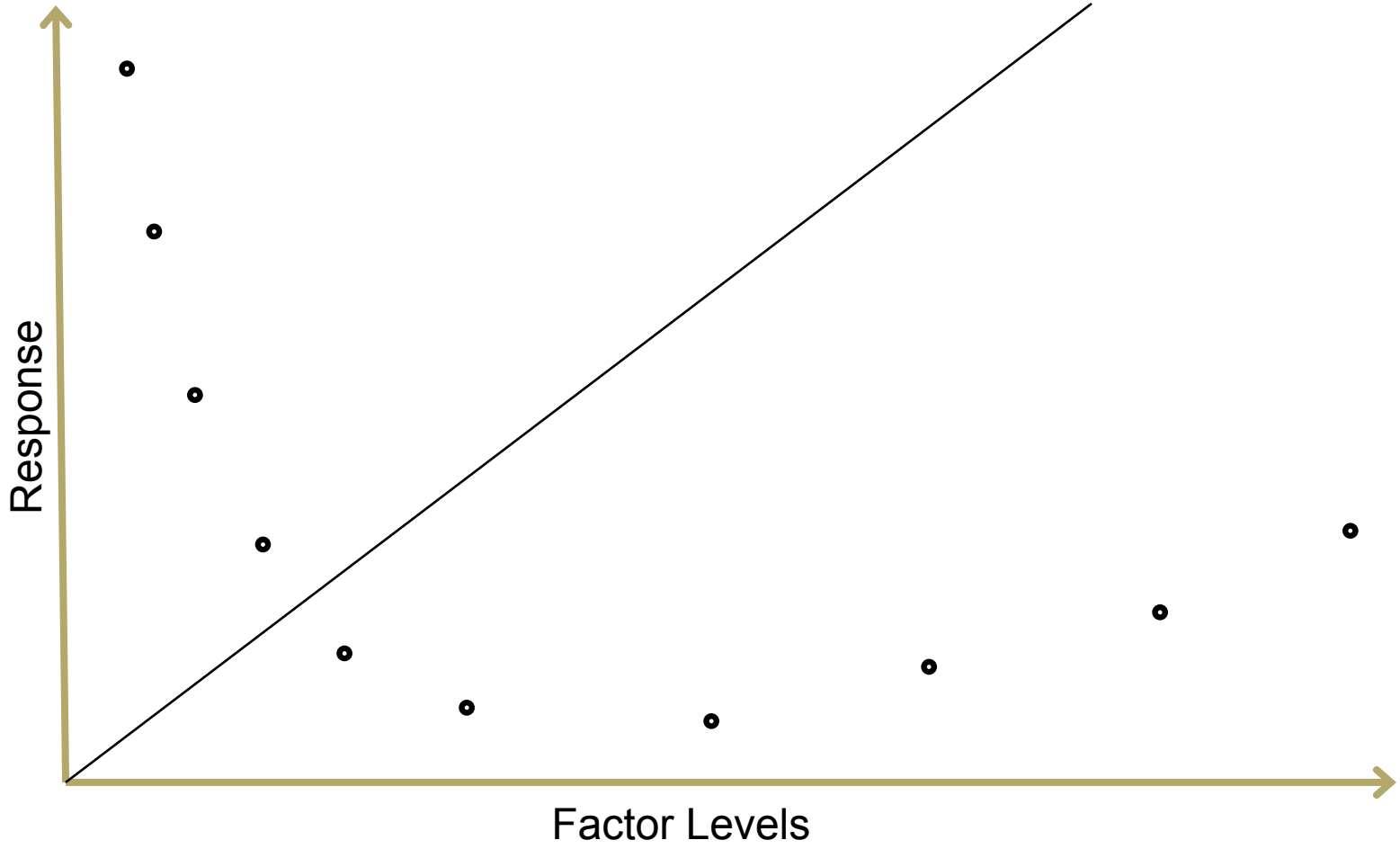
Saddles

- Quadrant saddle: interaction terms twist the paper



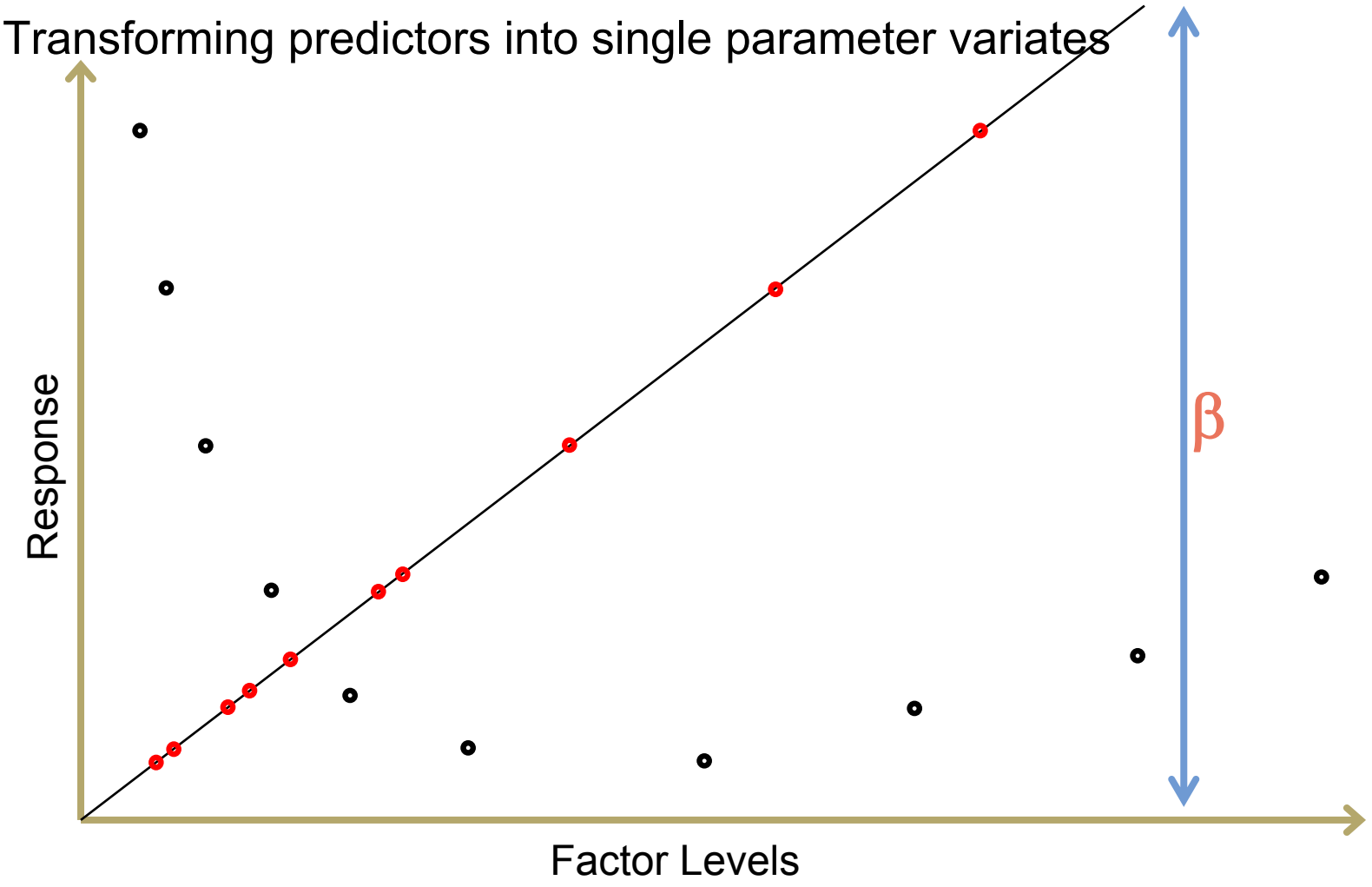
Saddles

- Transforming predictors into single parameter variates



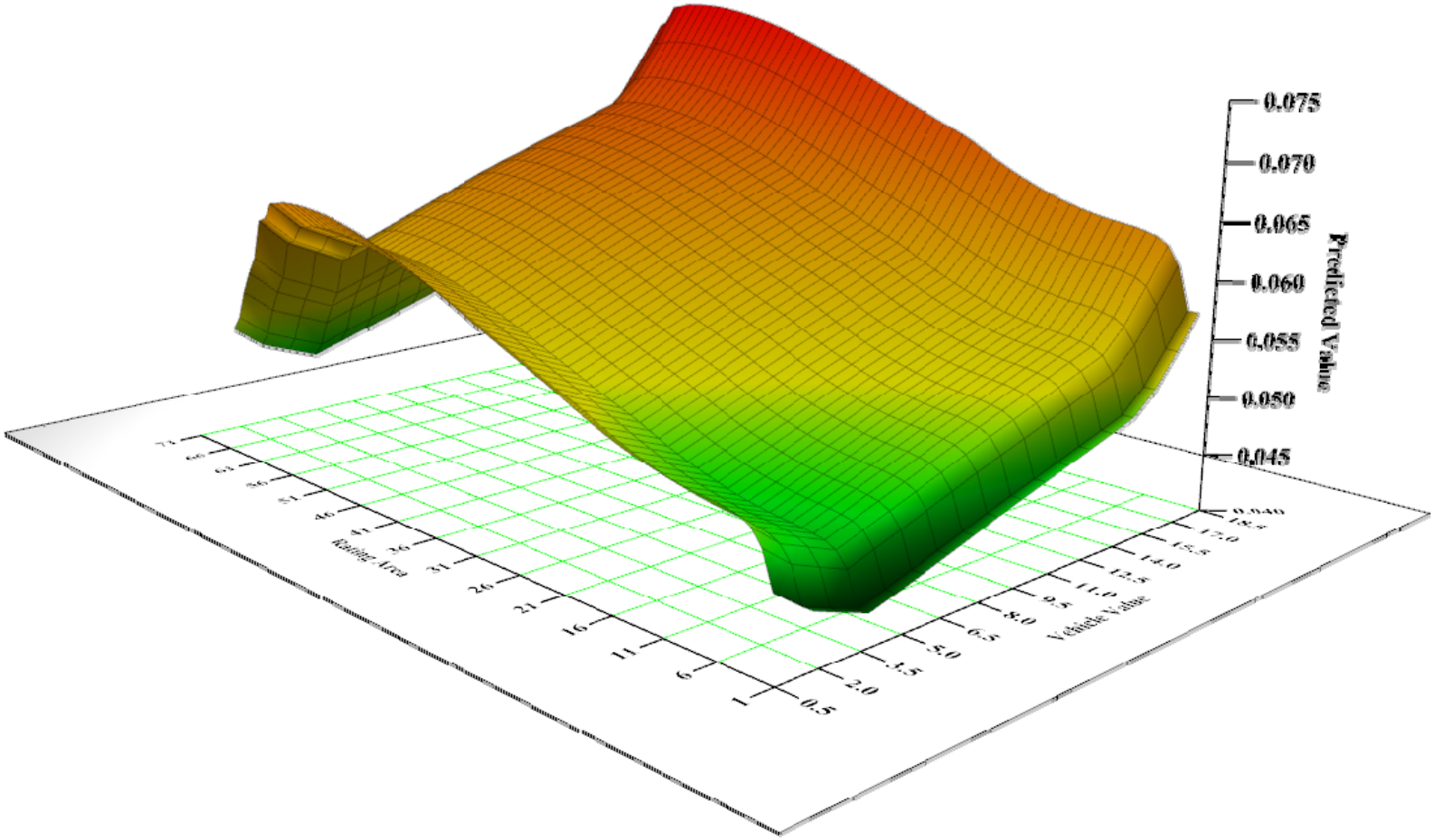
Saddles

- Transforming predictors into single parameter variates



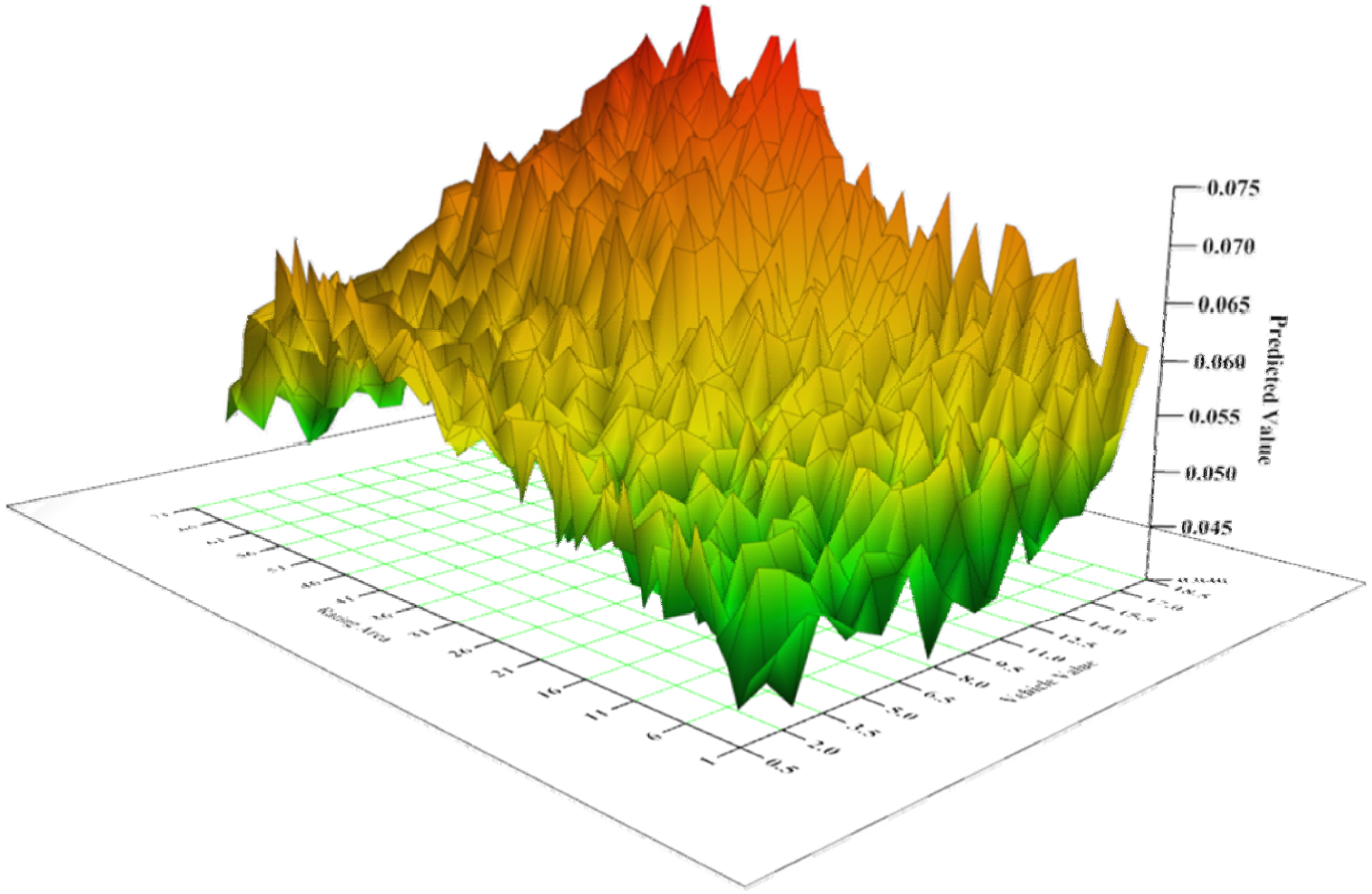
Saddles

- Case study: vehicle value x rating area



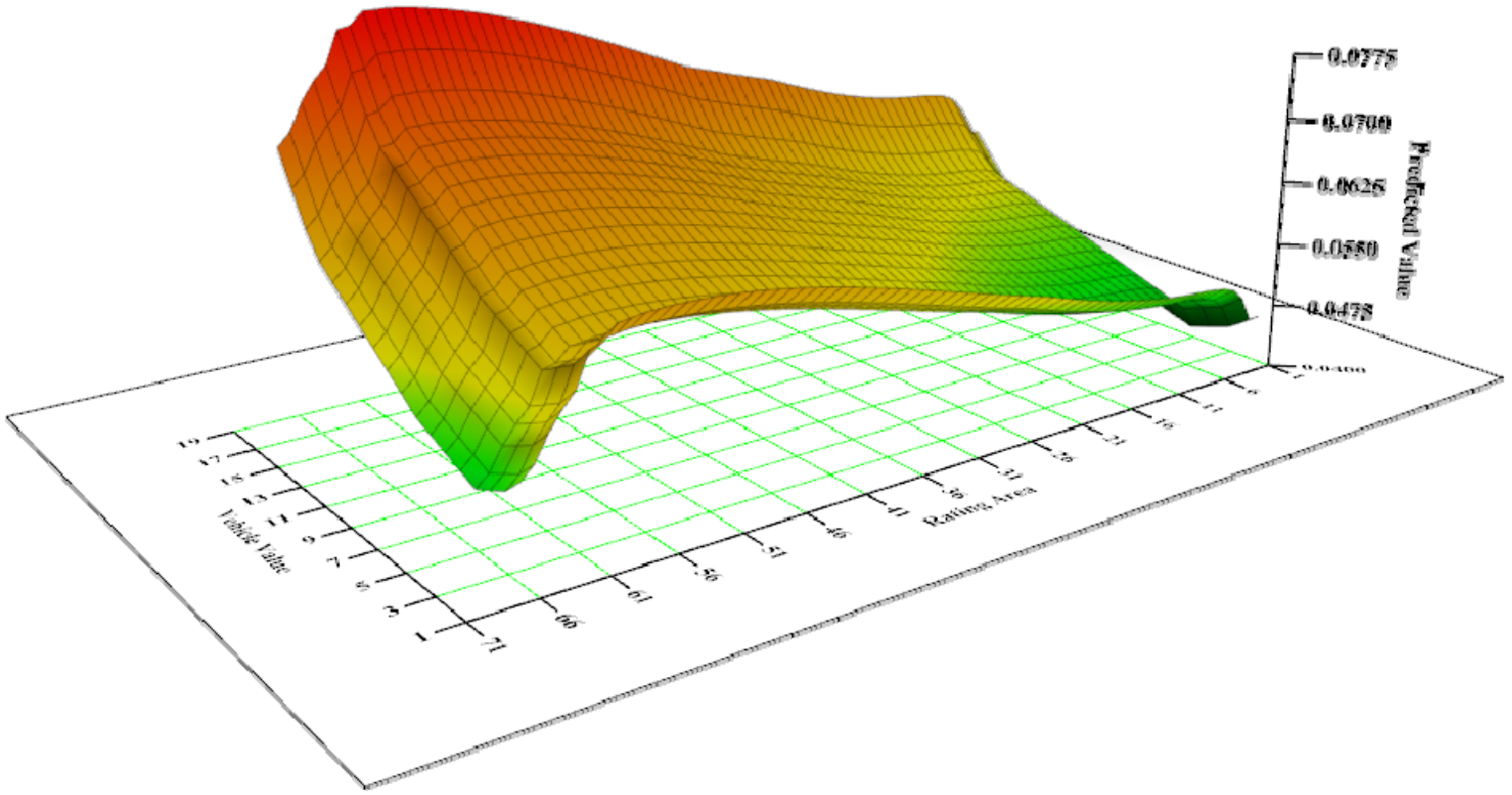
Saddles

- Full interaction is very noisy



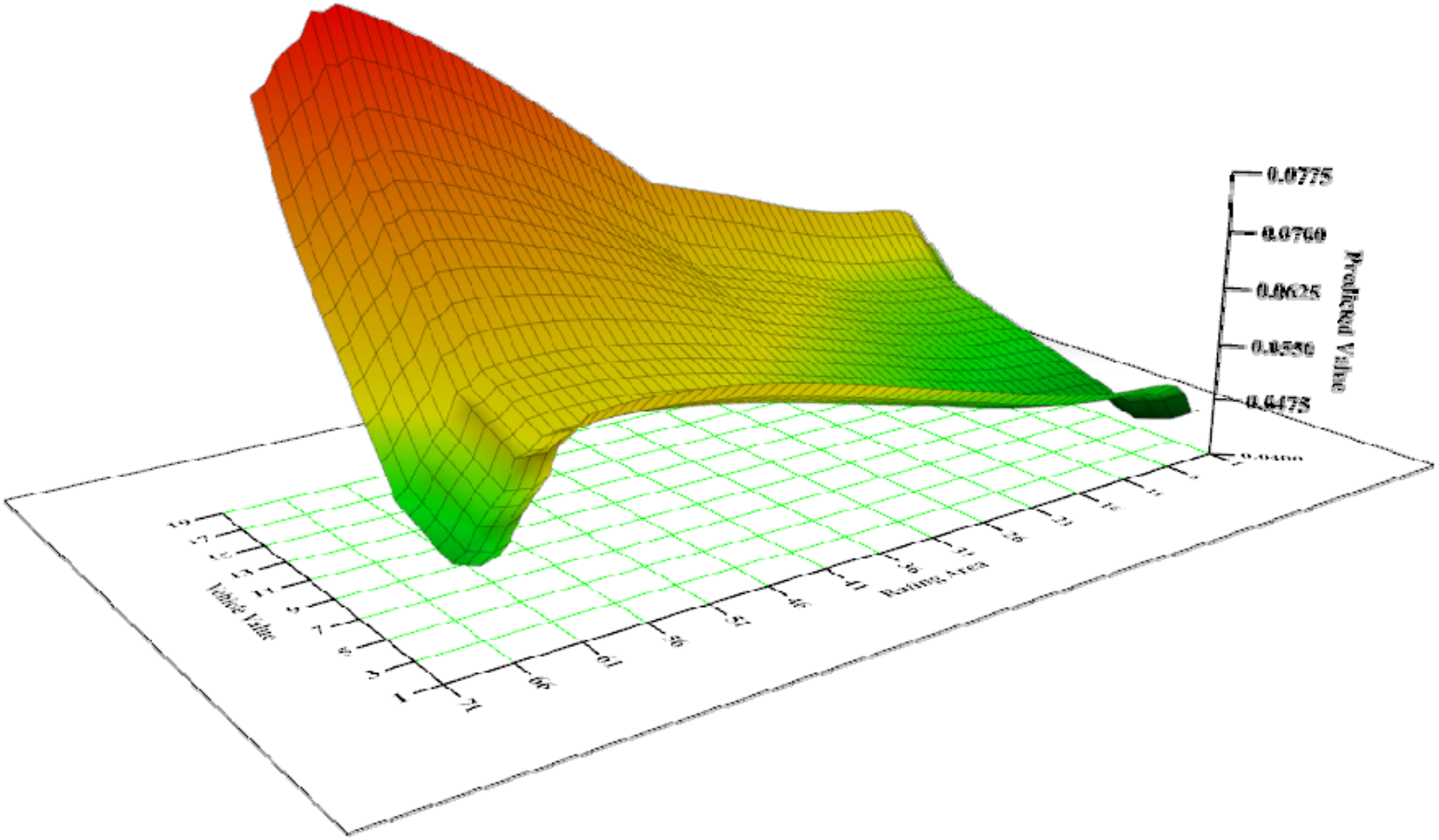
Saddles

- Different quadrants to be tested



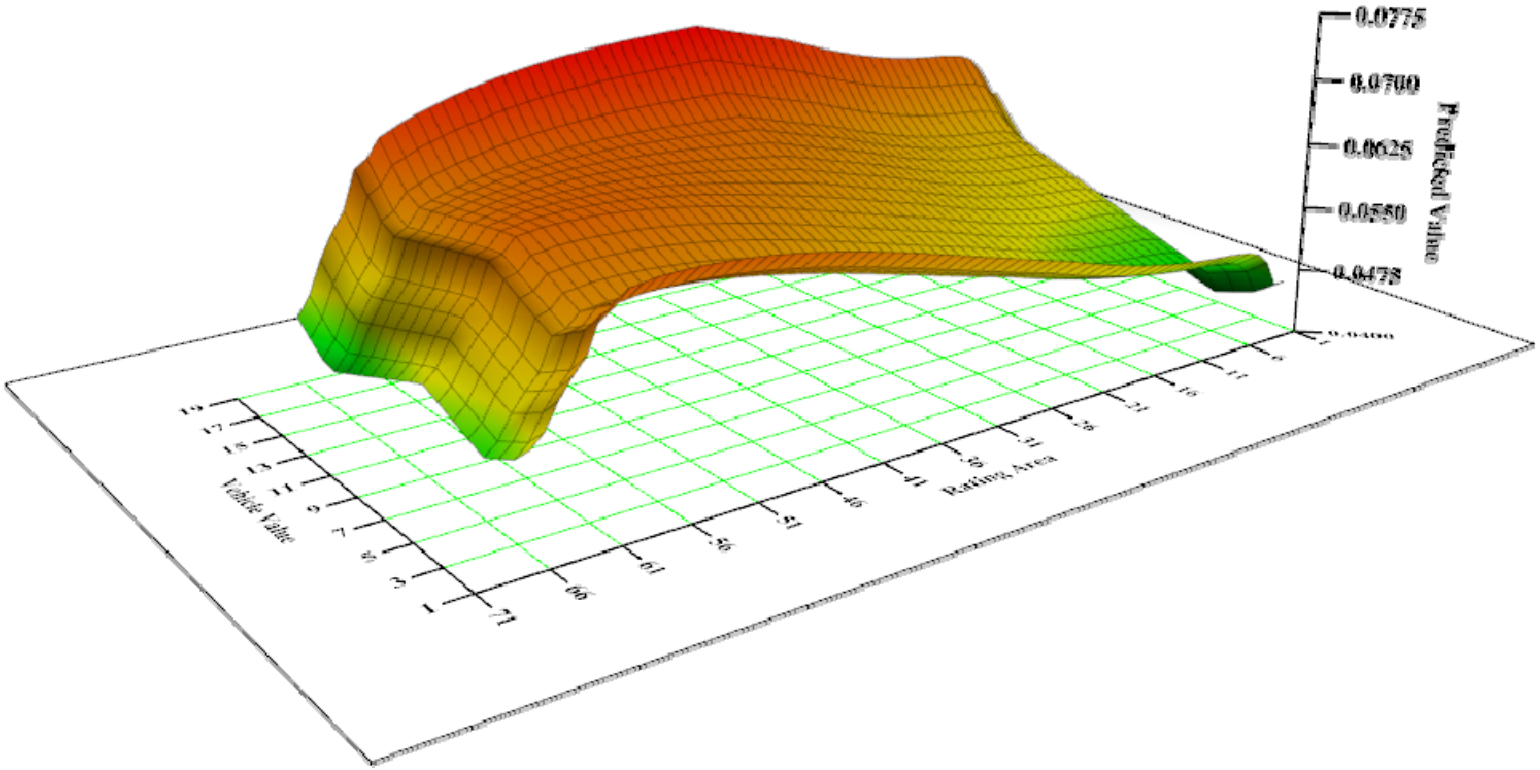
Saddles

- Focus on higher valued vehicles in certain areas



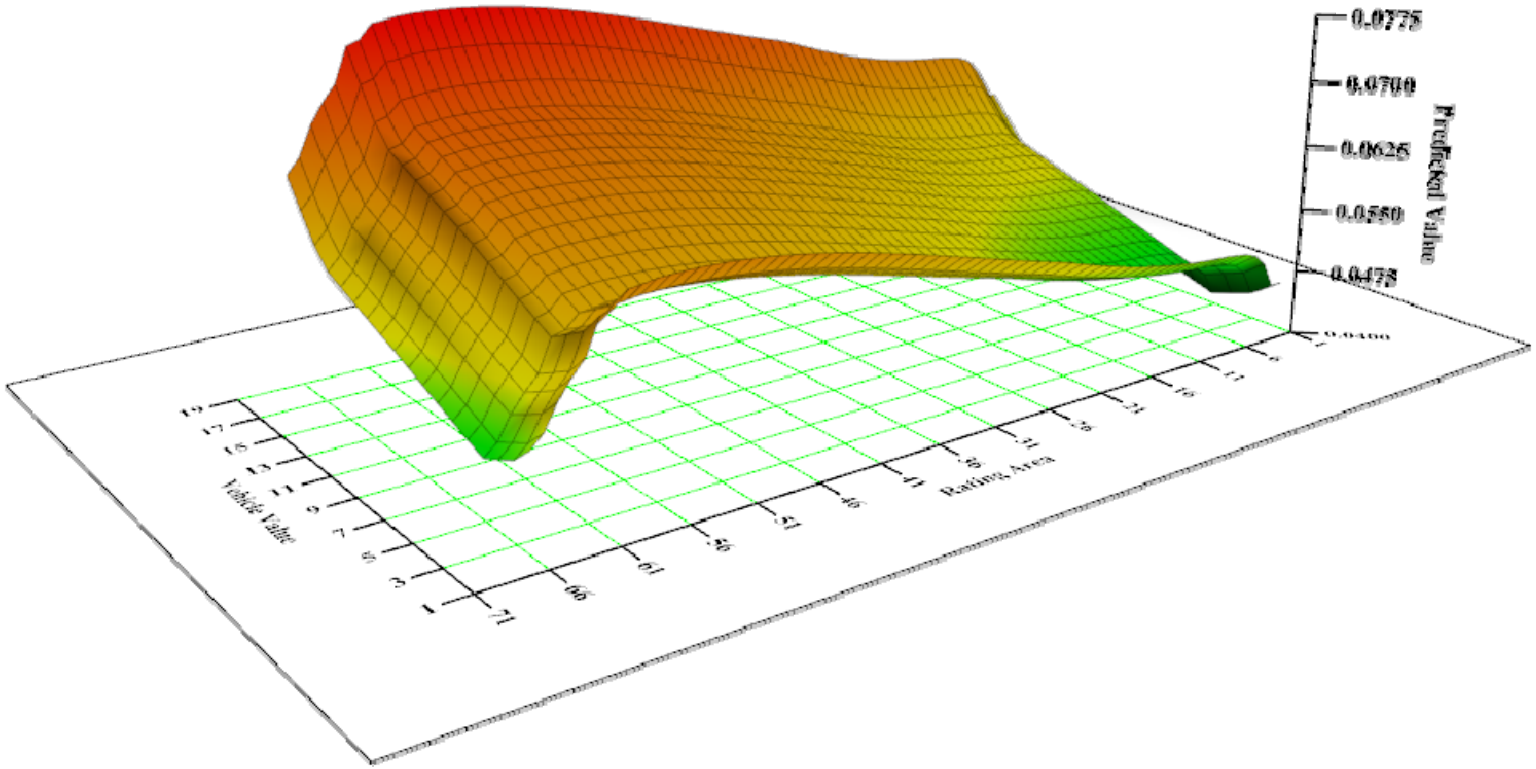
Saddles

- Systematically study different twists



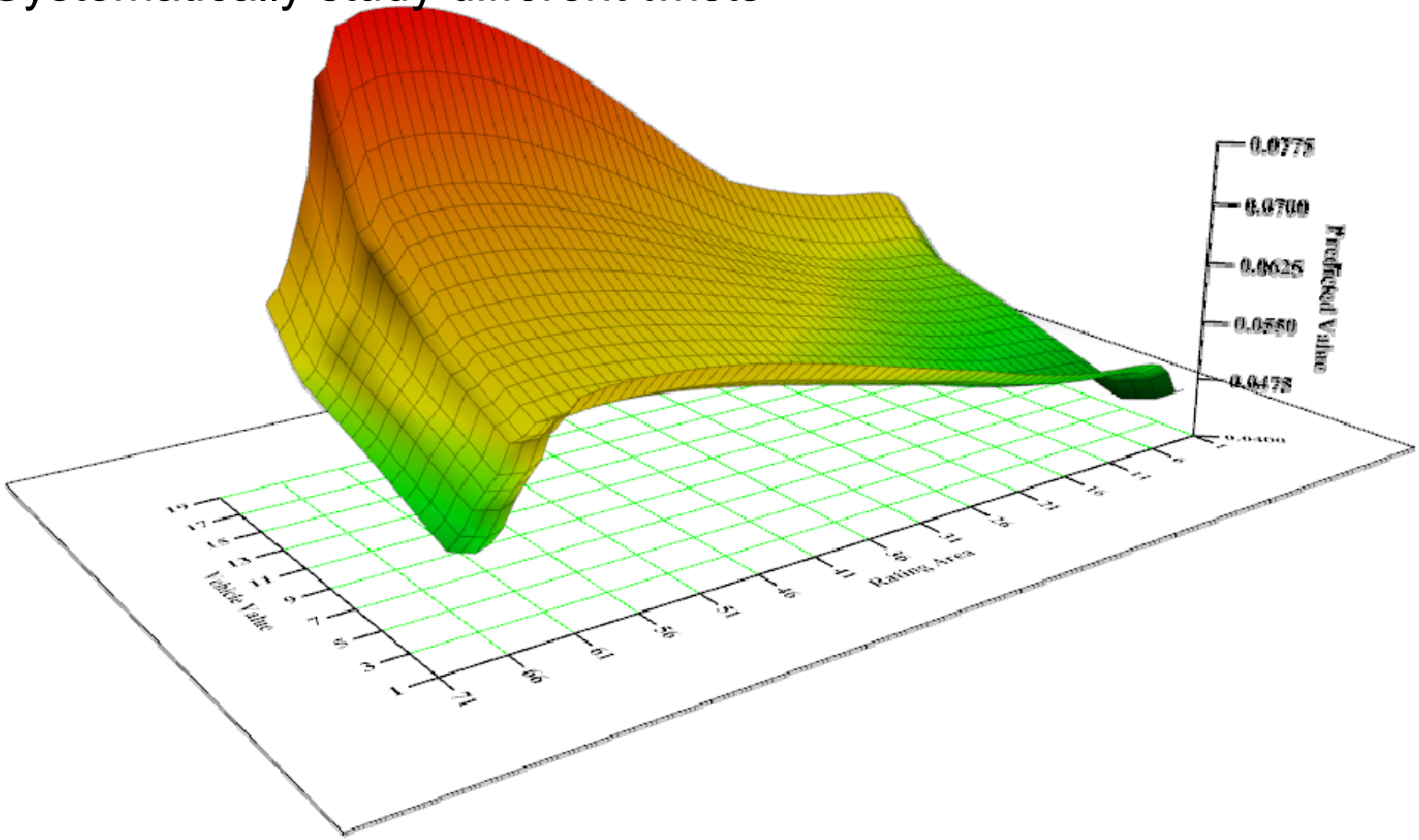
Saddles

- Systematically study different twists



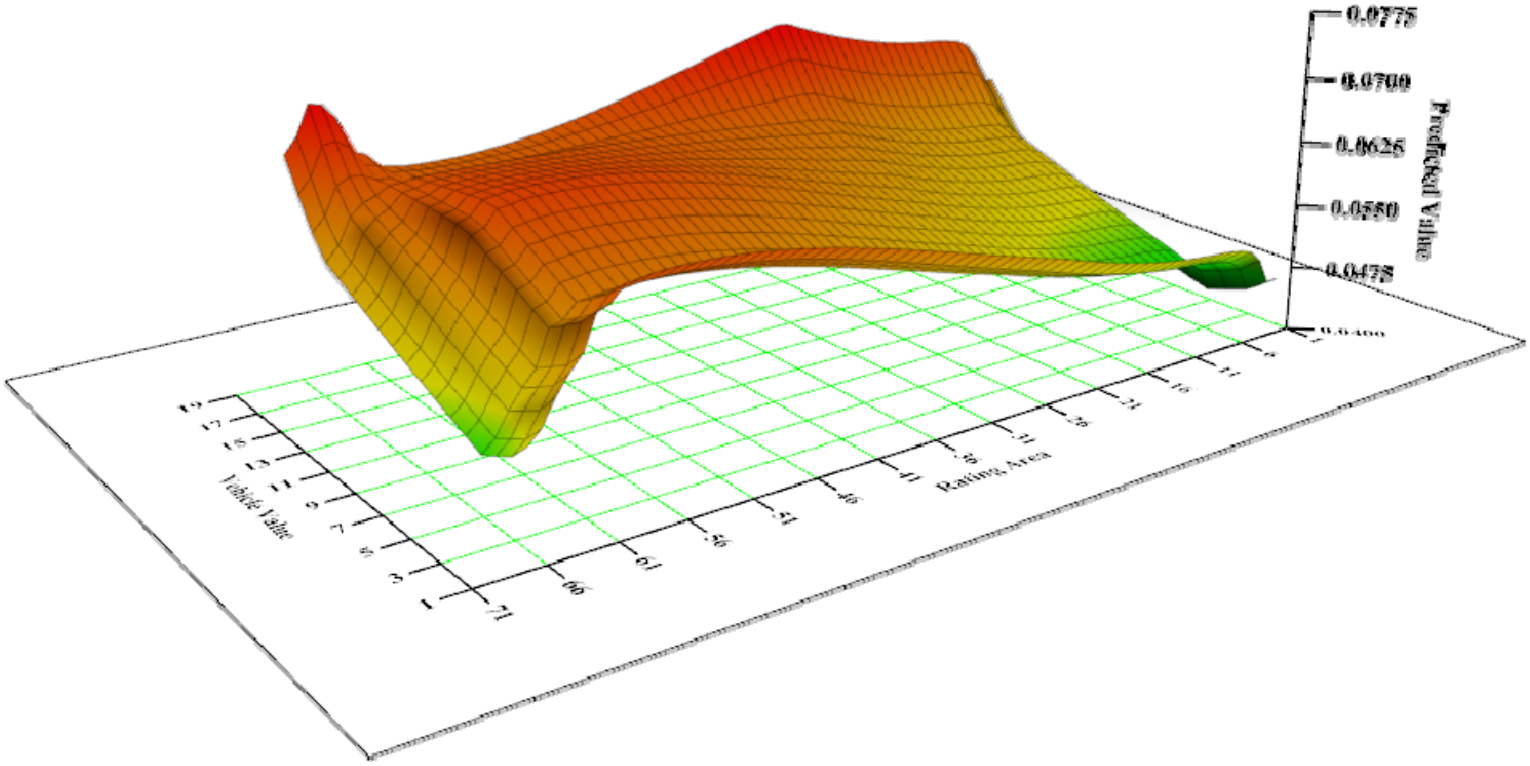
Saddles

- Systematically study different twists



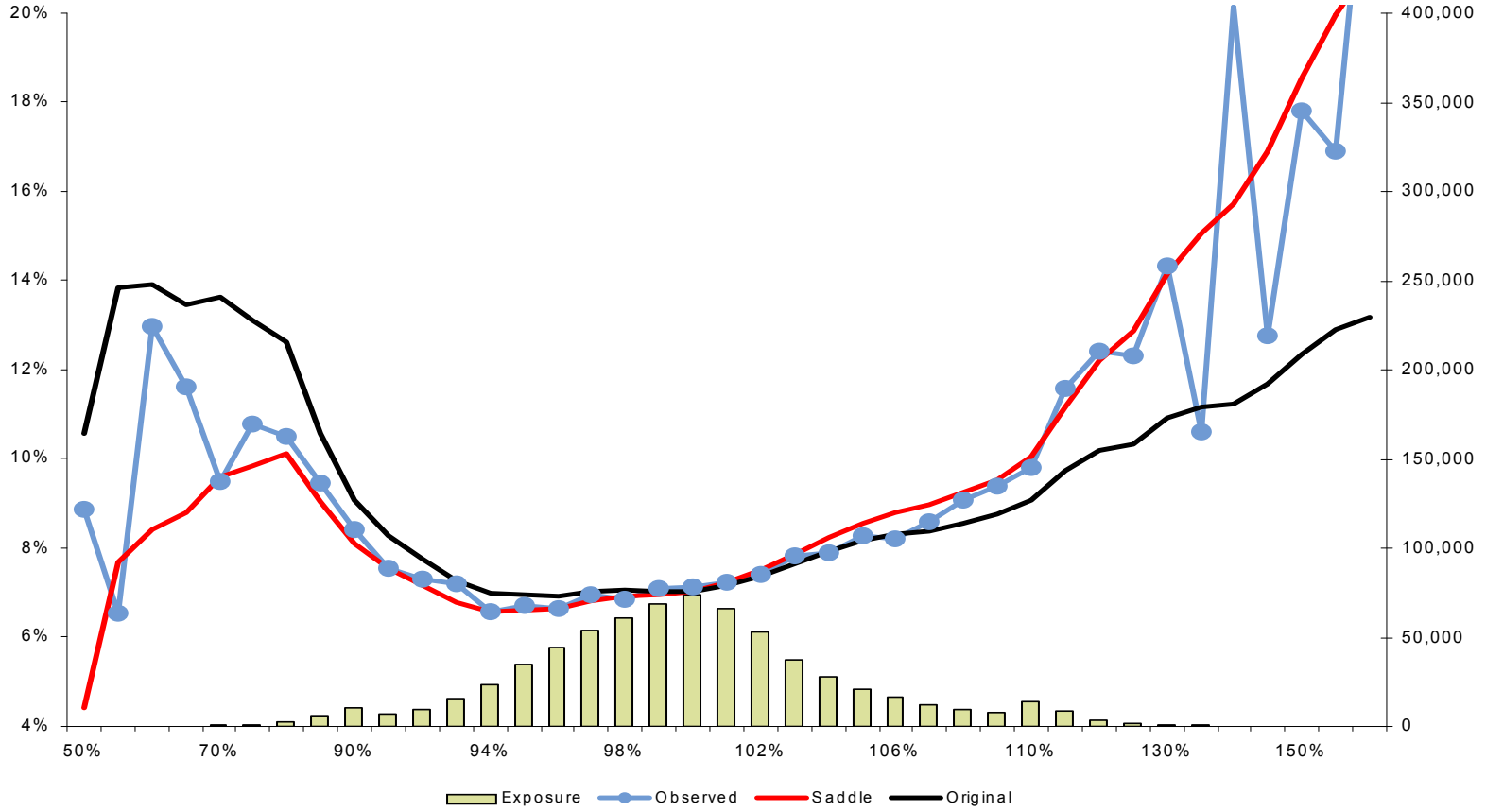
Saddles

- Systematically study different twists



Saddles

- Frequency out of time sample



Conclusion

- Interactions are an important part of the model creation process
- Volume of data requires a process to systematically study interactions
  - Balance testing
  - Decision tree tools
  - Saddles
- Effort needed in simplifying and validating identified interaction
  - Saddles use the framework of variate vectors in the design matrix to quickly simplify and validate new interaction terms