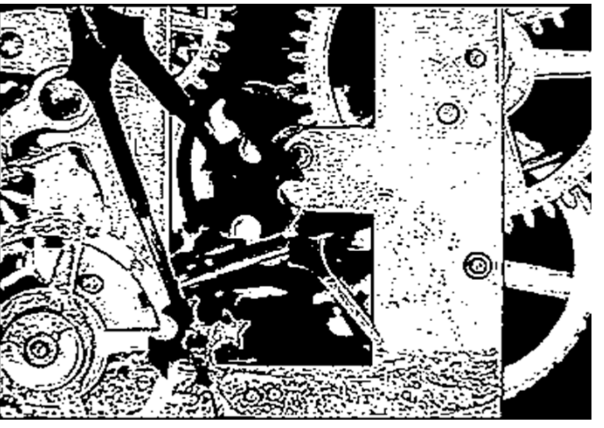
Partial least squares regression (PLS) path models

Predicting multiple target variables
Overcoming strong correlations



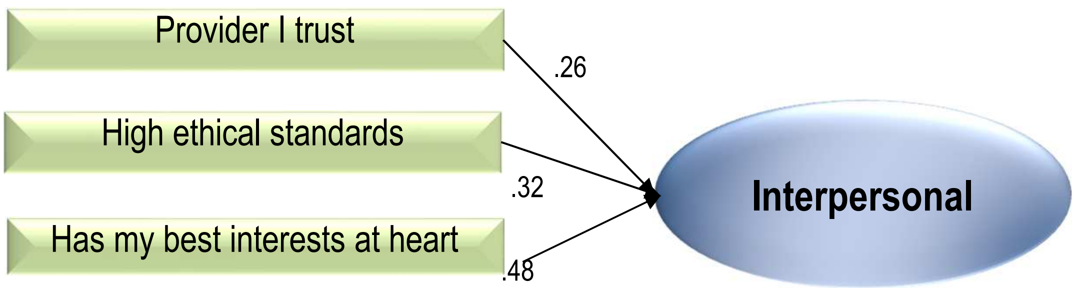
Steven Struhl

Converge Analytic

Predict two or more targets; deal with strong correlations

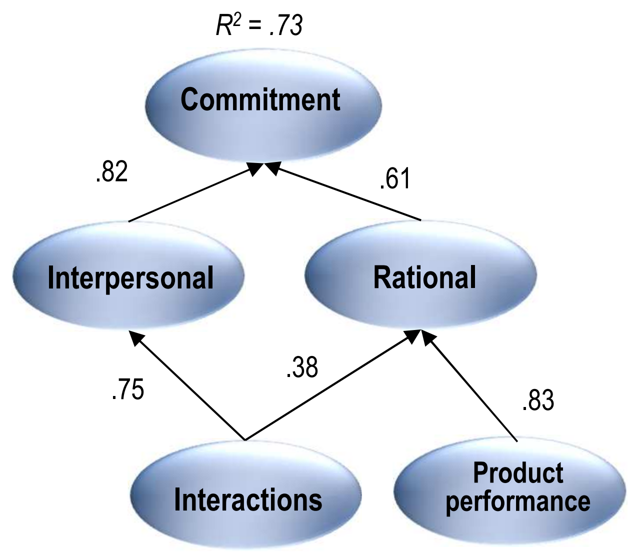
- Partial least squares (PLS) path models greatly extend standard regression
  - You can predict multiple target (dependent) variables
  - Highly correlated variables pose no problem
    - They get grouped into sets like the factors in factor analysis

Three highly related variables formed into a group that we have named "Interpersonal"



Variable groupings lead to strong prediction

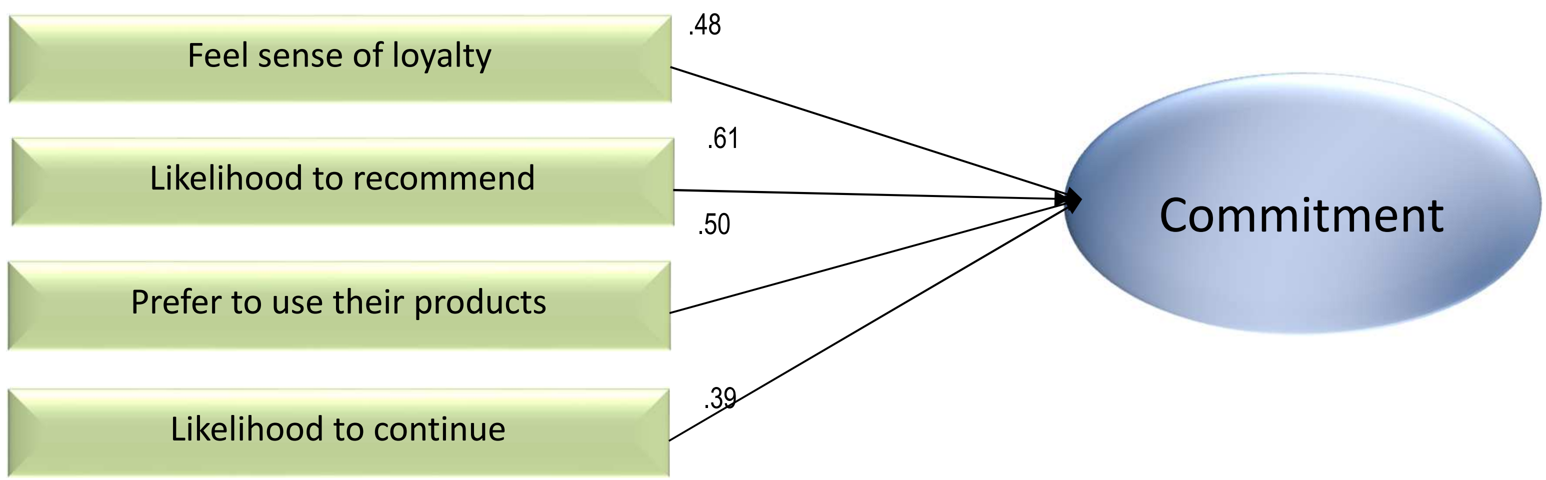
- Structures in the data are revealed by the **paths** between sets of variables and the target variable(s)
  - **Paths** also may exist among these sets
- These paths tell a story about what drives the target variable
- Here we have just the variable groups—and their strengths
  - Strengths range between 0 and 1
- **Interpersonal** has more effect on the target **Commitment** than does **Rational**
- Some of the groups go into others
  - **Product performance** goes into **Rational**
  - **Interactions** goes into **Rational and Interpersonal**
- Prediction of **Commitment** is very strong—R^2 is 0.73



Converge Analytic

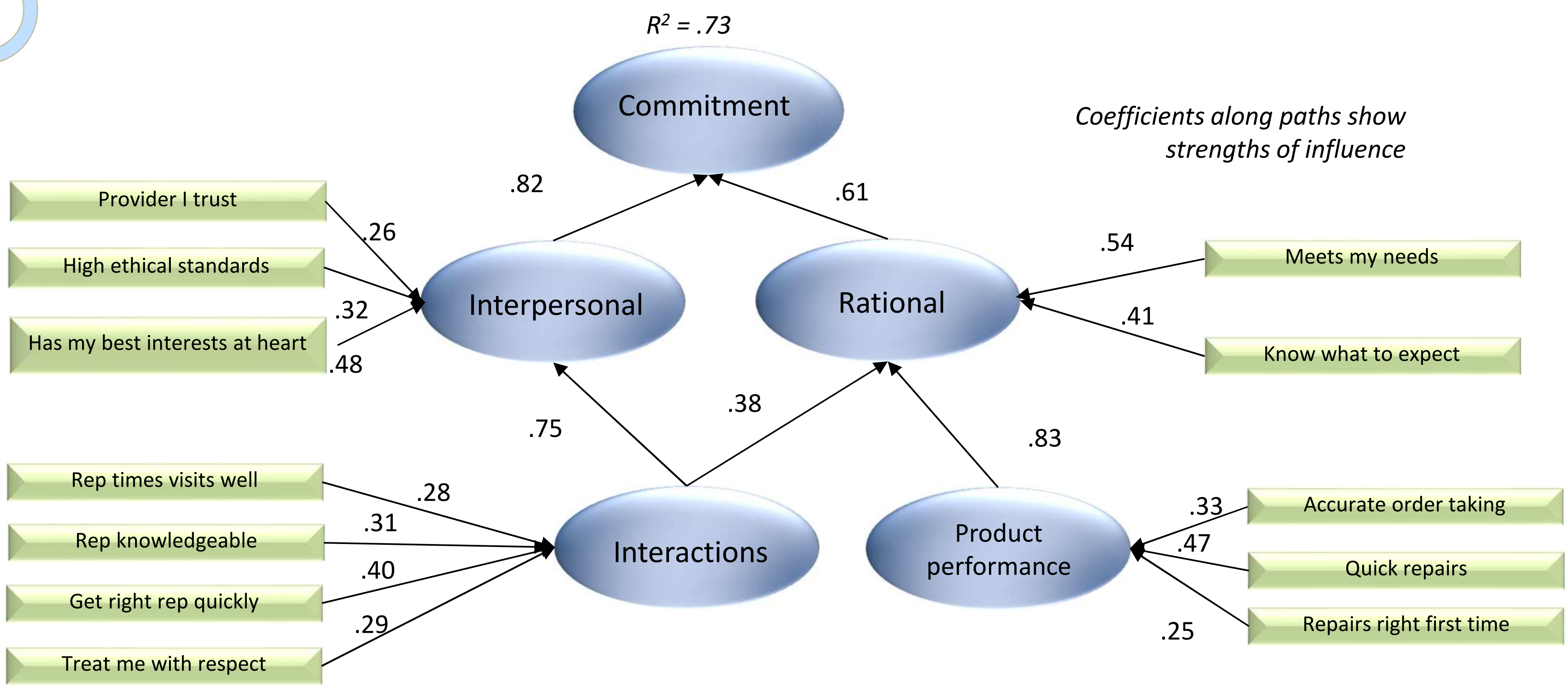
Target combines four separate variables

- The target is four separate variables combined
- We see the relative strengths of these variables
  - Likelihood to recommend is the strongest component by a small margin with a weight of 0.61
  - Likelihood to continue is the weakest by a small margin at 0.39
- Weights under 0.10 usually are not significant when tested



The whole model with the variables feeding into the groups

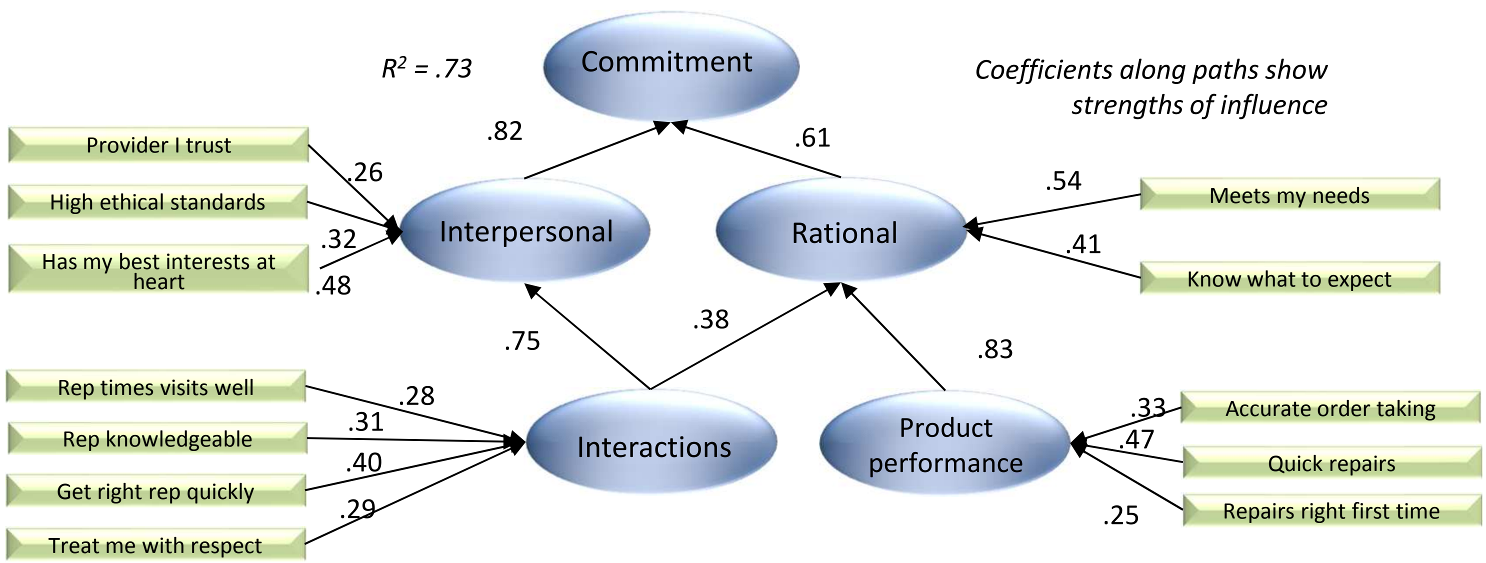
- This is the whole picture—first a chance to look at it—explanation follows on the next page.



Converge Analytic

The whole model with the variables feeding into the groups

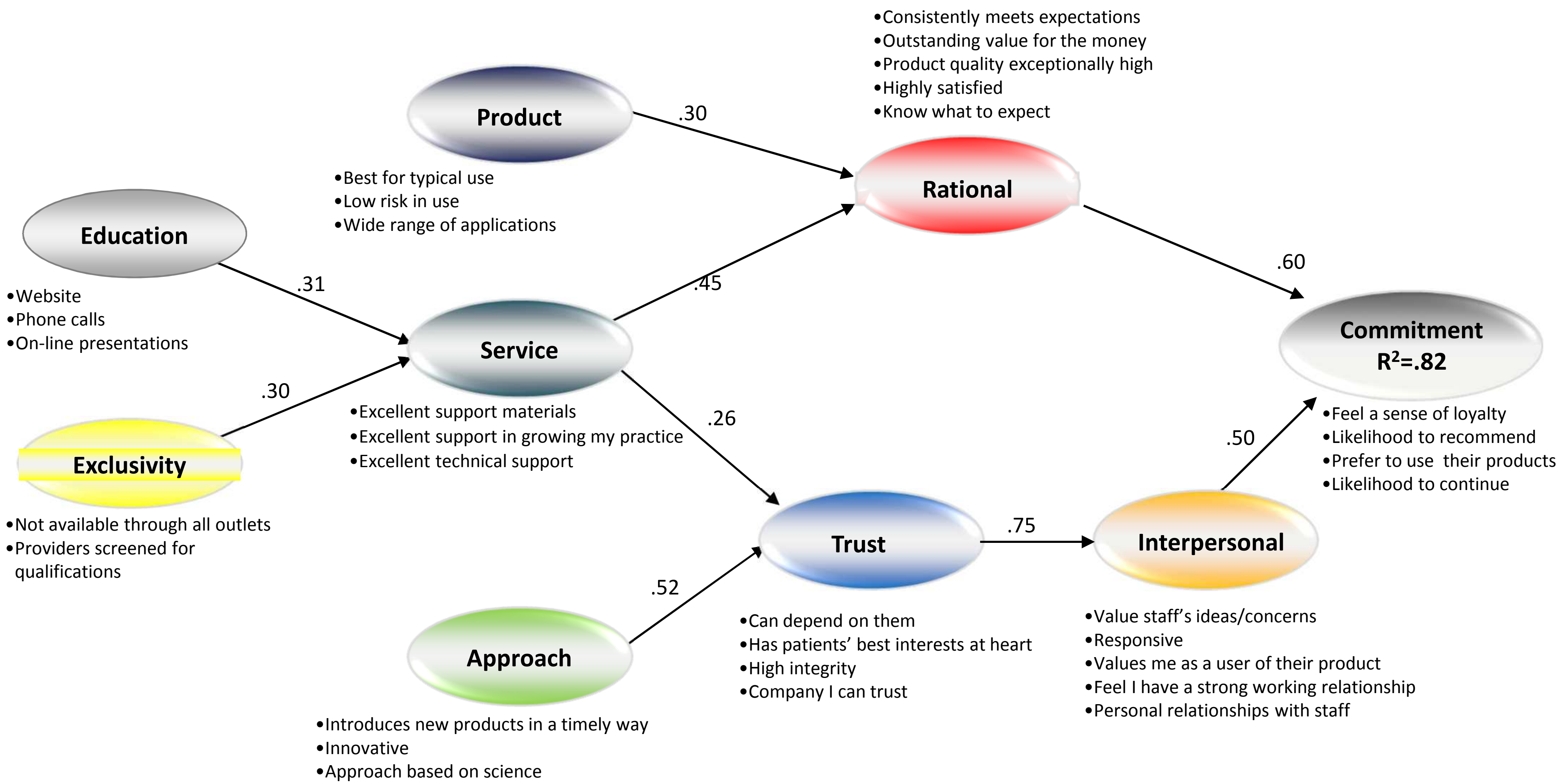
- **Interpersonal** has a stronger influence on **Commitment** (coefficient .82) than **Rational** (coefficient = 0.61)
- Some variables influence just the **Interpersonal** group (e.g., **provider I trust**, etc., shown to the upper left)
- Some variables influence only the **Rational** group (e.g., **meets my needs**, and **know what to expect**, shown to the upper right)
- Variables in **Interactions** (lower right) influence both **Interpersonal** and **Rational**
- Variables in **Product performance** dimension influence only the **Rational** side
- Commitment is explained very well—the R^2 is very high at 0.73



Converge Analytic

PLS handles many variables and complex models

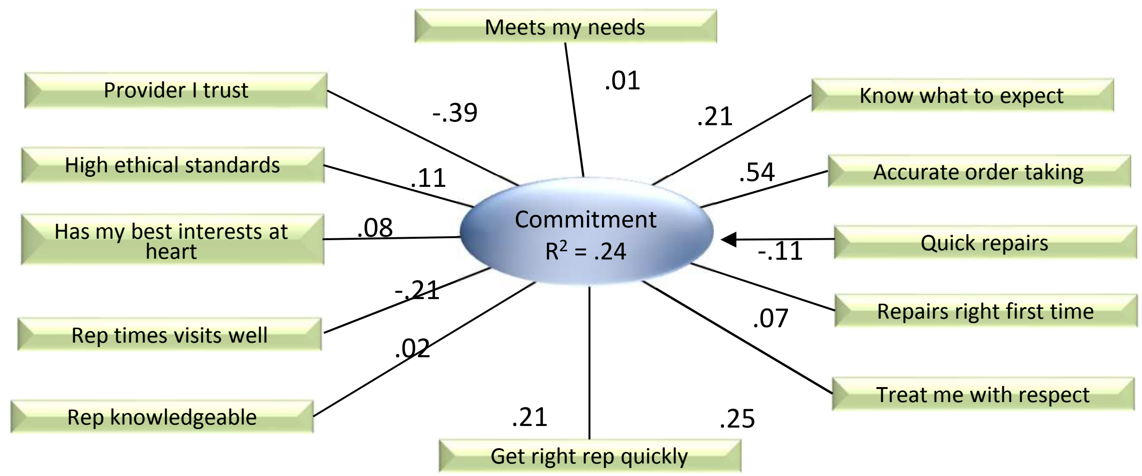
- PLS path models can handle models with a great many variables
- Coefficients going into the groups are not shown for ease of reading

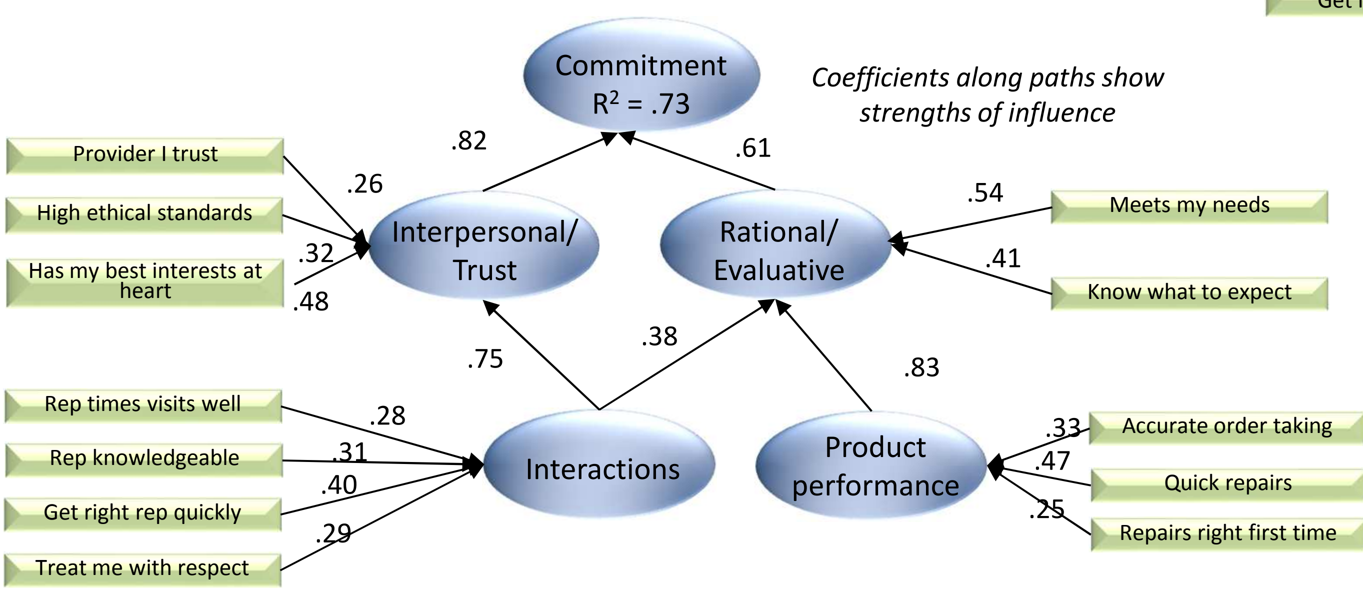


PLS path model does better than standard regression

The same variables in a PLS model and standard regression—standard model shows weakness

- Standard regression: Everything goes the same way to the same end point
- No ideas are identified
- Coefficients get distorted and reversed
- Low R^2 for target variable



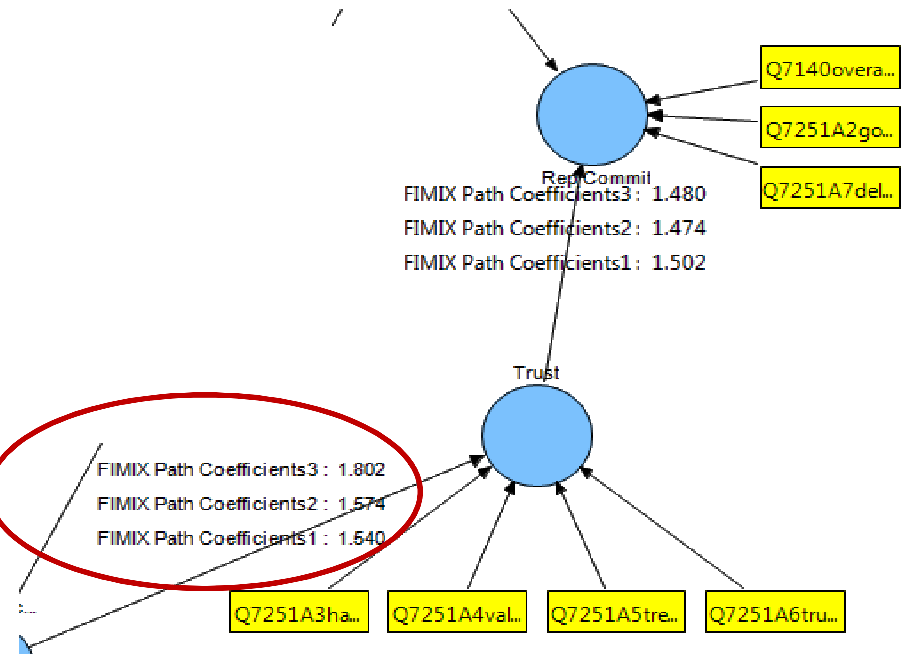


Coefficients along paths show strengths of influence

- PLS model shows relationships
- We see how strongly basic groups or ideas influence the dependent, as well as the strengths of the individual variables measured

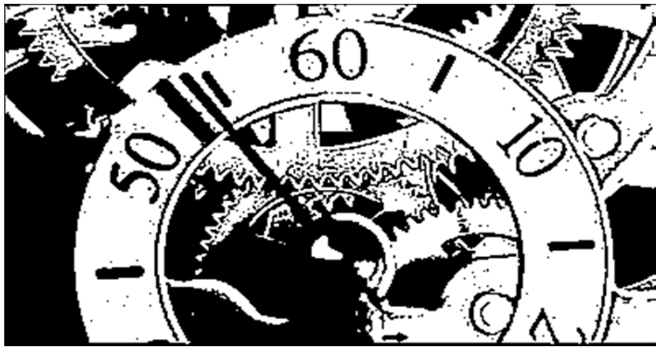
Extensions to PLS bring new analytical power

- Several powerful extensions to PLS have brought it still further beyond traditiona regression models
- This shows a segmentation based directly on the PLS model, using a method called “Finite Mixture Models” (FIMIX)
  - The groups have different coefficients, showing directly what is more and less important to each
  - Every respondent has a likelihood of belonging to each segment
- Other capabilities are being developed
  - For instance, **moderator effects** can show interactions among the variable groups



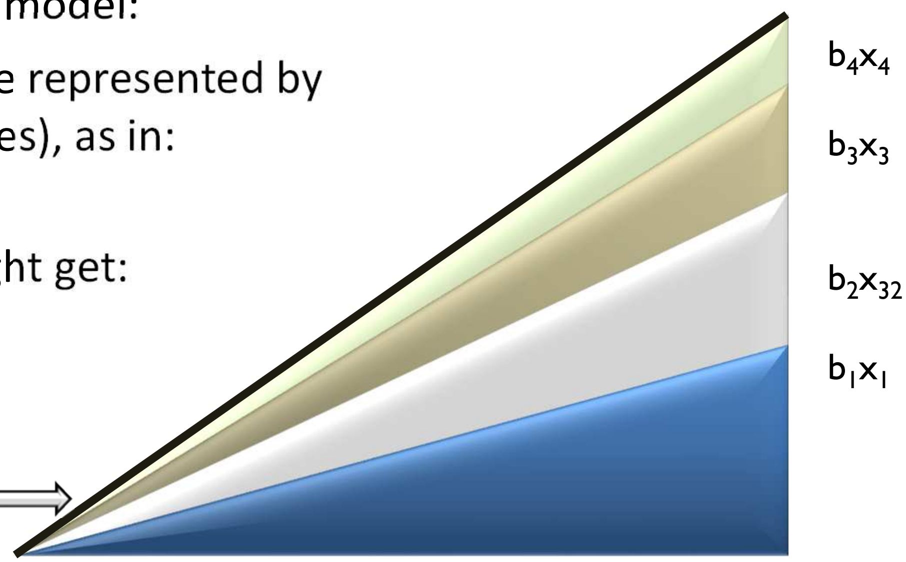
Appendix
In depth comparisons and key references

Standard regression
Structural equation models
References



Standard regression is like making a mixture

- Standard regression (**ordinary least squares**)[1] starts with this view—
  - The target variable (**y** or dependent variable) forms a straight line, and
  - If you add the predictor variables in the right proportions, this mix will sum to the value of the dependent
- Variables do not get grouped and no themes emerge
- Standard regression creates a familiar model:
  - The proportions of the variables are represented by different values (shown as "b" values), as in:
    $b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 = y$
  - Filling that in with numbers we might get:
    - 0.6x1 + 0.4x2+ 0.2x3 + 0.15x4 = y



This is the value of the dependent along this line. We get it by summing the predictor variables after we multiply each by its coefficient

There is more of b_1 than any of the others in this mix and less of b_4. The value of the dependent (the heavy line) goes up perfectly as you increase the other variables, something we often can only wish to see in real life

[1] Also called OLS regression

PLS vs. SEM—differences in basis concepts

- PLS differs from the (covariance-based) SEM approach in underlying conceptualization
  - PLS analyzes causes and make predictions in highly complex situations
    - Data drives the model as much as theory
    - It works well for the applications and predictions typically used in research
  - The structural equation modeling approach is more theory-oriented, and may seek to confirm or deny a theorized set of relationships among variables
    - See, for instance, Anderson and Gerbing, 1988.
- With PLS, sample sizes can be smaller in SEM analysis
- Some authorities consider PLS better suited for explaining complex relationships
  - *PLS comes to the fore in larger models, when the importance shifts from individual variables and parameters to packages of variables. . .in large, complex models with latent variables PLS is virtually without competition* –Wold, 1985
    - Also see Fornell, Lorange, and Roos, 1990

PLS advantages over SEM

- Compared with SEM, PLS has a strong advantage in that does not make strong demands (or assumptions) about what the data must be like so it can be analyzed
  - It returns useful results without requiring very specific measurement scales, sample sizes, or error distributions
- PLS also avoids two particular problems that can cause analyses to stop when using the SEM approach
  - These problems are known as “inadmissible solutions” and “factor indeterminacy.”
    - For discussions, please see Fornell and Bookstein, 1982
- PLS is under active development and has several strong new capabilities, such as segmentation based directly on the model and new methods of model testing and refinement

Key references

- Anderson, J.C. and Gerbing, D.W. (1988). "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103(3), 411-423.
- Fornell, C., and Bookstein, F. (1982). "Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory," *Journal of Marketing Research*, 19, 440-452.
- Fornell, C., Lorange, P., and Roos, J. (1990). "The Cooperative Venture Formation Process: A Latent Variable Structural Modeling Approach," *Management Science*, 36(10), 1246-1255.
- Wold, H. (1981). "The Fix-Point Approach to Interdependent Systems: Review and Current Outlook," in H. Wold (Ed.), *The Fix-Point Approach to Interdependent Systems*, Amsterdam: North-Holland, 1-35.
- Wold, H. (1985). "Partial Least Squares," in S. Kotz and N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences (Vol. 6)*, New York: Wiley, 581-591.

Questions? Comments?

Dr. Steven Struhl

smstruhl@convergeanalytic.com

smstruhl@gmail.com

☎ 847-624-2268