Following contours:
Parameter stability as an inference tool in models of decision-making

Sabina J. Sloman Daniel Oppenheimer

Carnegie Mellon University

Can we develop interpretable tests of computational models?

As the field of judgment and decision-making progresses, more and more computational models are introduced to explain choice phenomena.

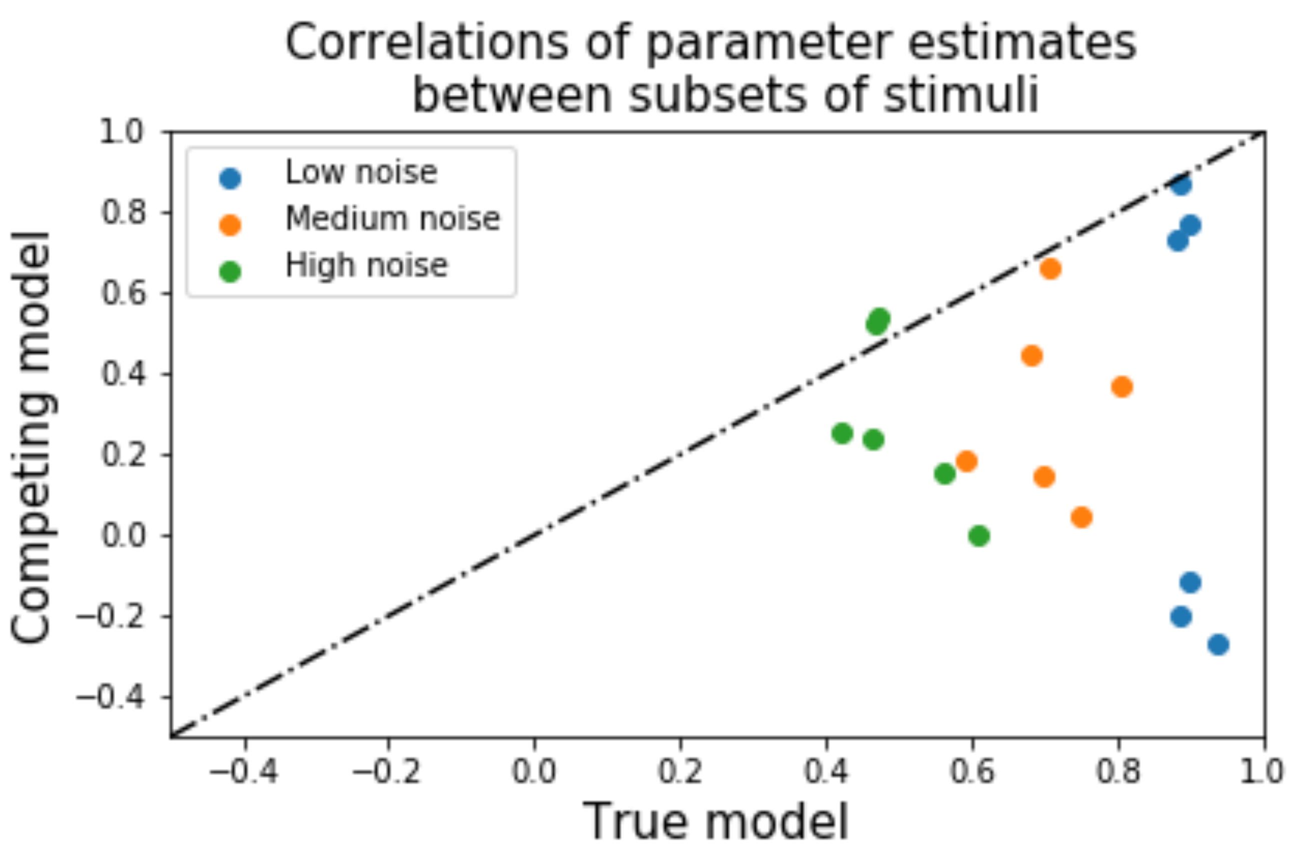
Quantitative model tests are common, but difficult to connect to theorized behavioral patterns.

We connect a quantitative property of a model that can be used for model selection to theorized behavioral patterns.

Method: We arbitrarily choose a model to simulate data from (a voting agent model of preferences [1]). We refer to this model as the "true" model. We then analyze the behavior of decision makers (DMs) under the "true" model. We also analyze DMs' behavior under a "competing" model, a version of the pairwise normalization model [2].

Quantitative measures of within-participant parameter stability recover the true model.

Computational models retain information from behavioral data in the form of parameter estimates. Quantitative methods that estimate out-of-sample fit are premised on the assumption that the true model can most effectively transfer information from DMs' choices on one set of stimuli to their choices on other stimuli. The amount of transferable information can be measured by the degree of stability in estimated parameter values.



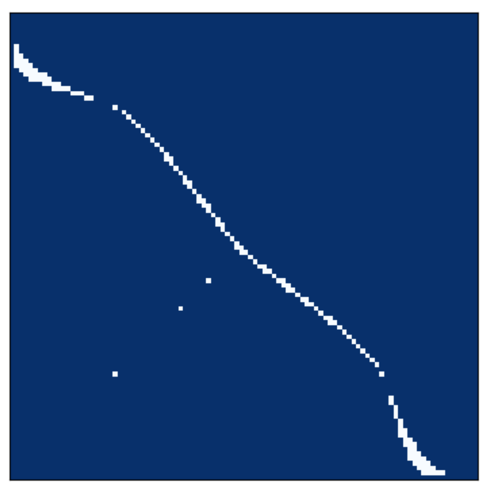
[1] Anouk S. Bergner, Daniel M. Oppenheimer, and Greg Detre. VAMP (Voting Agent Model of Preferences): A computational model of individual multi-attribute choice. *Cognition*, 192:103971, November 2019.

[2] Peter Landry and Ryan Webb. Pairwise Normalization: A Neuroeconomic Theory of Multi-Attribute Choice. *SSRN Electronic Journal*, 2019.

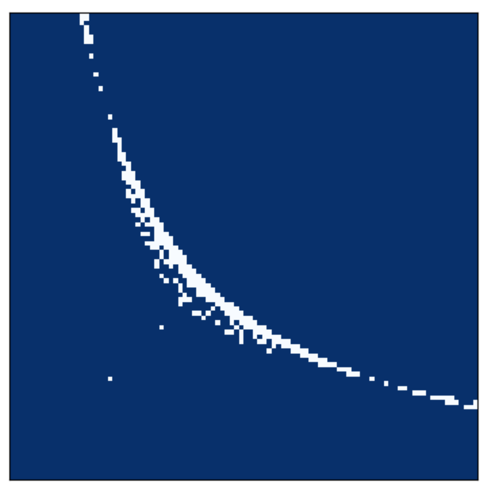
Parameter contours visualize the behavioral predictions of different parameter settings.

Parameter contour: A set of observations that discriminates between adjacent parameter values

For a parameter k, we identify where adjacent values make discriminating predictions...

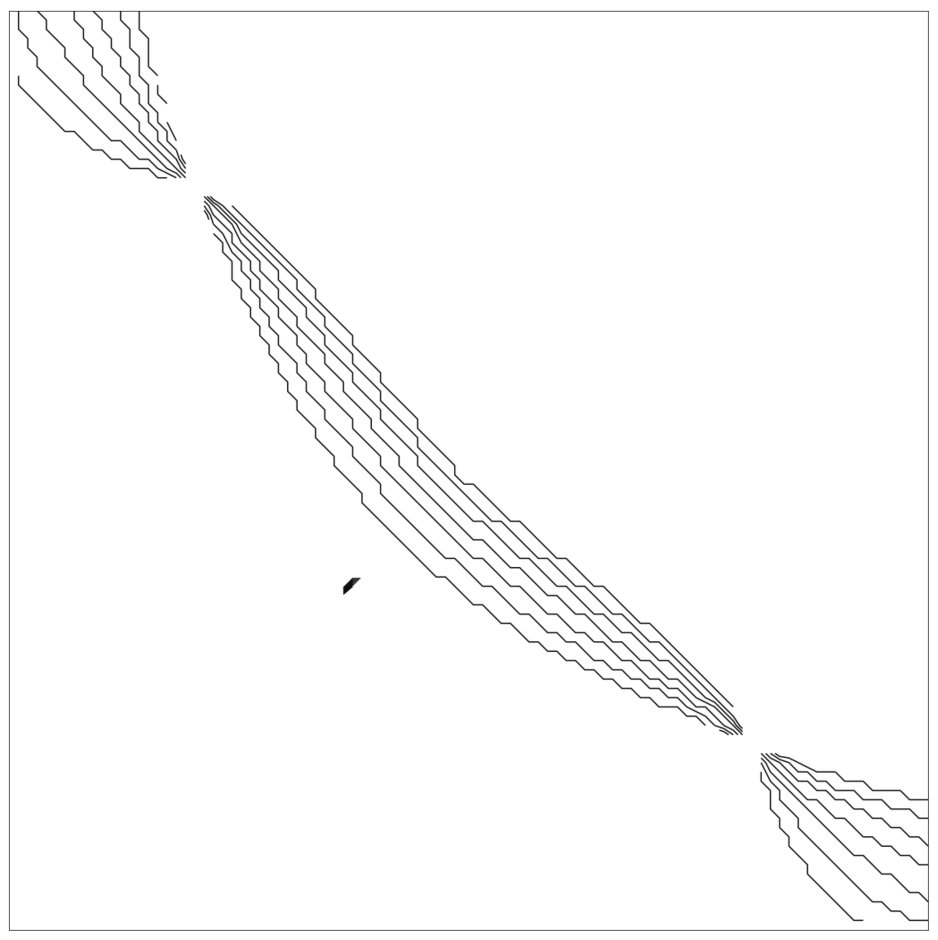


$k = .1$ vs. $k = .15$



$k = .45$ vs. $k = .55$

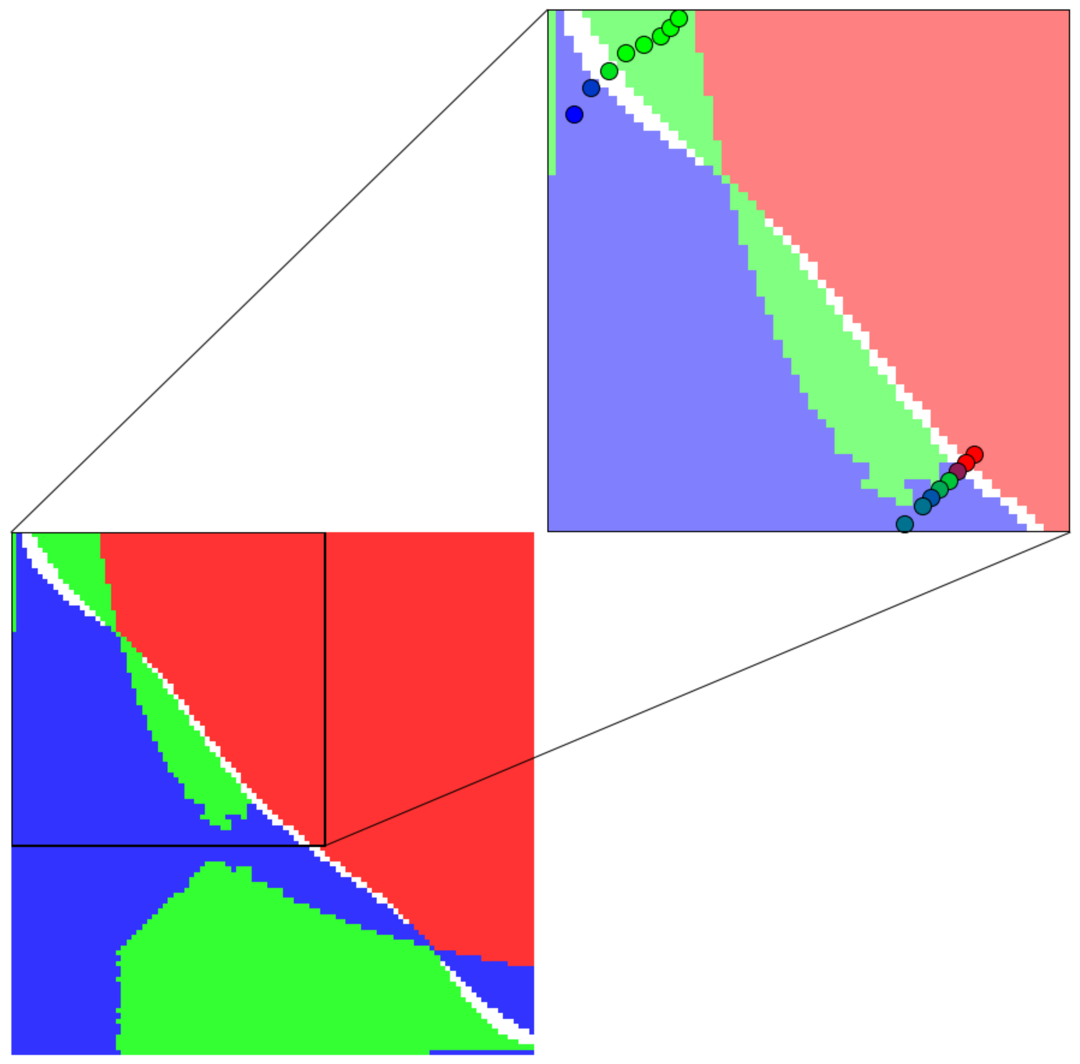
...and stack these predictions into a full set of **parameter contours**.



Parameter stability implies certain behavioral patterns across contours.

Experimental stimuli (dots) can be selected to fall within parameter contours. The model can make predictions of the choices of a DM with any $k = k^*$ (background colors). The model predicts that choices will "switch" at the contour corresponding to $k = k^*$ (white).

This corresponds to a behavioral prediction of the form "DMs who choose blue on this stimulus will choose red on that stimulus"... which will only hold *iff* the DMs' parameters are stably estimated as k^*.



Why? The colors of the dots show the choices of DMs whose k was estimated to be k^* using data from the upper half of these stimuli. Since the model is right, it successfully predicts their choice pattern on the lower half of stimuli. Because these stimuli straddle the parameter contours, this pattern again recovers the DMs' k as k^*.

Takeaway: Understanding how changes in parameter values relate to changes in behavior facilitates interpretable model testing.

- The relative stability of the parameter estimates of two models can recover the true model.
- Parameters encode transferable information about a participant's behavior.
- This same information can be represented as predicted choice patterns among carefully-selected stimuli.