

Quo vadis Robustness?

S. Morgenthaler^{1*}

¹ *École polytechnique fédérale de Lausanne; stephan.morgenthaler@epfl.ch.*

**Presenting author*

Keywords. *Model uncertainty; Bias; Regression, Big data.*

1 Exploratory Data Analysis v. Robustness

Robustness provides an answer to model uncertainty. It has dealt successfully with distributional uncertainty in linear models and with multivariate observations, in particular through theoretical advances and resulting heuristics. But the emphasis on asymptotic theory and a narrow view of model uncertainty has held back the general use of robust procedures.

The more archaic concepts favored by JW Tukey such as sensitivity and resistance should again take precedence over asymptotic robustness. I will give examples from John Tukey's teaching materials to illustrate his thinking. He put the focus on generally applicable ideas – or technologies – that could contribute to the construction of procedures for data analysis. Examples include the use of weights, transformations, the jackknife, and borrowing strength. In any given problem, a combination of technologies can be applied, with each combination giving different results. The emphasis is not on achieving optimality, but rather on creating a variety of data analytic procedures, which could enter into competition and be compared by simulation and through examples.

This does not argue for or against robust procedures. It argues for variety and creativity. The problem it poses is rather one of dissemination. We need a collection of the basic concepts that we find useful and instructions on how to use them. We also need a depository for all the robust/resistant procedures that have been found, for the results that are known about them, and for comparisons among themselves and with other approaches.

2 New Avenues for Robustness

The basic principles of robustness can lead to new procedures in many data analytic problems where it is unclear how to think about alternatives to established stochastic models. What is a robust/resistant analysis of risk or of extreme observations? What is a robust/resistant allocation of resources in the face of uncertainty. How would a robust/resistant test of genetic risk look like? What is a robust classification.

I will use a simple situation involving binary observations to illustrate problems of bias that arise naturally in many of these applied areas and comment on the use of the basic principles in big data applications, where it is natural to repeatedly fit subsamples and to combine the results.