

## Drifts, Shifts and Instabilities to Quantify Scorecard Model Risk.

### Abstract

This talk explores how Data Drift changes scorecard specification. Data Drift is usually considered a danger for models, but this presentation takes a different view: by taking control of Data Draft and assessing the Model Specification Drift under well-chosen data perturbations, we quantify Model Risk effectively.

This view emerges from the elegant theory of Statistical Geometry [1]. First we formulate a first-order method that converts directly a perturbation of development data into a perturbation of model parameters. Second we use second-order structure such as curvature to quantify model robustness.

This talk describes how this helps manage models and quantify the model's stability and specification risk:

- as a correction to Gerard Scallan's classic Delta method, accurate when scorecard data is highly interactive, achieved by replacing the Delta method's contravariant coordinate system with covariant coordinates;
- as a compensator of recognised deficiencies in SMOTE and of other methods that rebalance data;
- as a quantitative test of stability for data pre-processing methods such as Weights of Evidence;
- as a batch quantification of model validation challenges and "what if..." scenarios.

This approach can also be reversed: to find the ways in which model specification is most sensitive to Data Drifts. This leads to metrics designed to detect those data perturbations that cause most change to the model. Thus we augment Population Stability Index and other familiar blunt measurements of Data Drift, with high specificity monitoring of the most sensitive population changes. As an added advantage, this refined monitoring can be generated directly and automatically from the model.

Because this approach controls the Data Drifts actively and usefully, we prefer the term Data Shift. We propose this as a standard investigative tool in Model Risk quantification. To help this aim we have authored an R and Python library which we use to illustrate the analysis of this presentation.

[1] Amari, S., & Nagaoka, H. (2000). *Methods of Information Geometry*. Translations of Mathematical Monographs, Vol. 191. American Mathematical Society and Oxford University Press.

### Authors & Affiliations

Dr Alan Forrest<sup>1</sup>, Mr Andrija Djurovic<sup>2</sup>

<sup>1</sup>Credit Research Centre, University of Edinburgh Business School, Edinburgh, United Kingdom.

<sup>2</sup>Deloitte, Bucharest, Romania