The Annals of Applied Statistics
2017, Vol. 11, No. 4, 1901–1911
https://doi.org/10.1214/17-AOAS1041F
Main article: https://doi.org/10.1214/17-AOAS1041

REJOINDER: "ELICITABILITY AND BACKTESTING: PERSPECTIVES FOR BANKING REGULATION"

BY NATALIA NOLDE AND JOHANNA F. ZIEGEL

University of British Columbia and University of Bern

We would like to thank the discussants for interesting and insightful contributions. The discussants raised a number of diverse points, related to both theory underlying backtesting methodologies as well as to practical implications for banking regulation.

Robust traditional and comparative backtests. Chen Zhou clarifies the relation between the notion of identifiability of a risk measure and the ability to perform traditional backtests in the form of conditional calibration tests. We fully agree with him that in the absence of an identification function it is still possible to perform traditional backtests by assuming common properties of the conditional distributions across time. In our work, we have entirely focused on *robust* backtests as Zhou has phrased it, where robustness refers to robustness with respect to model uncertainty.

We would like to add that the same clarifications are in order for comparative backtests. Both elicitability and identifiability are only meaningful concepts when stated with respect to which class of distributions $\mathcal{P}$ they hold; cf. Definitions 1 and 2. Broadly speaking, the smaller the class $\mathcal{P}$, the weaker the condition for existence of an identification function or a strictly consistent scoring function for a given functional T. Let us give the following simple example: Suppose that $\mathcal{P}_s$ is a class of symmetric distributions. Then, for each $P \in \mathcal{P}_s$, the mean and the median coincide. Therefore, all consistent scoring functions for the median are also consistent scoring functions for the mean *relative to* $\mathcal{P}_s$, and the same holds for the respective identification functions. Relative to a class $\mathcal{P}_c$ of distribution functions such that all distributions have the same α-quantile, say $\mathrm{VaR}_\alpha(P) = c$ for all $P \in \mathcal{P}_c$, ES is identifiable and elicitable. Strictly consistent scoring functions can be obtained by setting $r_1 = c$ in equation (2.4). Similarly, the second component of the identification function at (2.7) with $r_1 = c$ identifies ES_α relative to $\mathcal{P}_c$. This is reflected in the ES backtest given by Zhou: The assumptions on the data-generating process allow to estimate c well enough that asymptotically we can work as if c was known.

Hajo Holzmann and Berhard Klar suggest comparative backtests for the entire tail of the P&L distribution instead of a specific risk measure; let us term them

Received August 2017.