The Markovian Information Criterion: theory, implementation and practical applications

Sylvain Barde

University of Kent

The recent increase in the breath of computational methodologies has been matched with a corresponding increase in the difficulty of comparing the relative explanatory power of models from different methodological lineages. In order to help address this problem a Markovian information criterion (MIC) is developed that is analogous to the Akaike information criterion (AIC) in its theoretical derivation and yet can be applied to any model able to generate simulated or predicted data, regardless of its methodology. Both the AIC and proposed MIC rely on the Kullback-Leibler (KL) distance between model predictions and real data as a measure of prediction accuracy. Instead of using the maximum likelihood approach like the AIC, the proposed MIC relies instead on the literal interpretation of the KL distance as the inefficiency of compressing real data using modelled probabilities, and therefore uses the output of a universal compression algorithm to obtain an estimate of the KL distance. Several Monte Carlo tests are carried out in order to (a) confirm the performance of the algorithm and (b) evaluate the ability of the MIC to identify the true data-generating process from a set of alternative models.

The criterion is then applied in a calibration exercise of a large-scale agent based model. Despite recent advances in bringing agent-based models to the data, the estimation or calibration of model parameters in such models remains a challenge. Most methods, such as the method of simulated moments (MSM), require in-the-loop simulation of new data. This may not be feasible for large-scale agent-based models, which typically are computationally heavy to simulate. This proposed application introduces an alternative 'large-scale' estimation approach for such models, applied to the Eurace@Unibi model. This model displays strong emergent behaviour and is able to generate a rich variety of nonlinear economic dynamics, including endogenous business- and financial cycles for certain parameter constellations. In addition, it is computationally heavy to simulate.

The method uses three stages. First, we use Nearly-Orthogonal Latin Hypercube sampling (NOLH) in order to generate a set of 513 parameter combinations with good space-filling properties. In the second stage we use the MIC to score the synthetic data against the empirical data. The third stage uses Kriging to build a surrogate model of the MIC, as a function of the parameters. The parameter combinations that provide the best fit to the data are then identified though optimisation of the surrogate model. Validation of the surrogate model is carried out by re-running the second stage of the analysis on the identified optimum and cross-checking the realised MIC against the MIC predicted by the surrogate model.