Editorial

2nd Special Issue on Applications of Optimization Heuristics to Estimation and Modelling Problems

The contents of this second special issue on applications of optimization heuristics to estimation and modelling problems reflect the broadening scope of successful applications of optimization heuristics in statistical analysis. The journal Computational Statistics and Data Analysis (CSDA) has acquired a positive reputation in publishing articles relying on results of optimization heuristics. This issue as the first one in 2004 (Winker and Gilli, 2004) aims to encourage further submissions in this field of research. It provides relevant examples for which optimization heuristics are found to work efficiently.

The fast increasing number of implementations, the introduction of new heuristics and the generation of a large number of hybrid methods results in a nontrivial selection problem. Given that most papers rely on a single method or compare just a small number of algorithms for a few problem instances, it is not obvious as to which method to choose for a specific application and problem instance. The classification of optimization heuristics provided in Winker and Gilli (2004) can help to identify crucial features of an algorithm. Nevertheless, its real performance on any given problem instance will remain subject to a careful implementation analysis. A recent overview on the methods and their potential can be found in Winker and Maringer (2007).

Typically, optimization heuristics are stochastic algorithms. Consequently, the results obtained for a specific problem instance might change with each repetition. Often, this stochastic feature is considered as being an unwelcome property, although statisticians are used to analyze stochastic events. In fact, the outcome of the implementation of an optimization heuristic can be analyzed using standard statistical methods as well as specific methods from Markov chain analysis. An examples for a statistical framework is presented by Fitzenberger and Winker (2007) in this special issue, while approaches based on Markov chain analysis can be found, e.g., in Jacobson and Yücesan (2004) and Jacobson et al. (2006).

Confidence in the methods will increase by an improved presentation of the implementations and a rigorous analysis of the results obtained. There is no commonly agreed standard for presenting results obtained from the application of optimization heuristics. For this special issue the following nonexhaustive guidelines have been considered:

- A complete description of the implementation of the heuristic including all parameter settings.
- The description of any preliminary testing of parameter settings (number of different settings analyzed, impact of parameter settings on outcome), if possible using a structured approach.
- A clear indication of how many runs of the algorithms have been used for each of the results presented. Furthermore, information on the distribution of the results (in terms of values of the objective function and/or other variables of interest) should be provided.
- Information whether and to which extent an increase in computational time (number of iterations etc.) improves the results should be provided. Preferably, some graph or estimate of the rate of improvement should be presented.
- The comparison with other methods will require either the evaluation of multiple problem instances or a Monte Carlo simulation. In both cases, a detailed information on the set up should be provided.

Now, a brief overview of the special issue contributions with regard to their application area and specific implementation details is provided. The first three papers deal with issues of model selection. Kapetanios (2007) uses Simulated Annealing and Genetic Algorithms for variable selection in econometric models. Thereby, he extends earlier approaches by Winker (1995) who used Threshold Accepting for lag structure selection in VAR-models. Hofmann et al. (2007) describe alternative optimization and heuristic strategies for deriving the best subset regression model which efficiently
avoids the enumeration of all subset models. The proposed strategies allow the tackling of large scale models. Ambrogi et al. (2007) use Genetic Algorithms for model selection in an artificial neural network setting.

The next three papers deal with issues related to clustering of data. Duczmal (2007) apply Genetic Algorithms for clustering spacial data allowing for irregularly shaped regions. Also Genetic Algorithms are used for determining stratum boundaries by Keskinturk and Er (2007). An application to bank rating systems is provided by Krink et al. (2007) who use differential evolution.

Applications to econometric modelling and estimation are the subject of the following three papers. An estimator for the censored quantile estimation problem based on threshold accepting is introduced by Fitzenberger and Winker (2007) who also provide a formal framework for the convergence analysis of estimators which are based on heuristic optimization procedures. Similar in spirit is the application to the estimation of nonlinear time series models presented by Yang et al. (2007), who use a hybrid of Genetic Algorithms and Simulated Annealing for this purpose. Staszewska (2007) describes how joint confidence intervals can be constructed for the impulse response paths of vector error correction models making use of heuristic optimization.

Finally, Ali (2007) describes how beta-distributions might be used to improve stochastic global optimization routines.

References


Manfred Gilli
Department of Econometrics,
University of Geneva and SFI,
40 Boulevard du Pont-d’Arve,
1211 Geneva 4, Switzerland
E-mail address: Manfred.Gilli@metri.unige.ch

Peter Winker
Department of Economics,
University of Giessen,
Licher Strasse 64, 35394 Giessen, Germany
E-mail address: Peter.Winker@wirtschaft.uni-giessen.de