Mathematical Statistics of Partially Identified Objects

Organised by
Victor Chernozhukov, Cambridge MA
Wolfgang Härdle, Berlin
Joel Horowitz, Evanston
Ya’acov Ritov, Jerusalem

21 April – 27 April 2013

Abstract. The workshop brought together leading experts in mathematical statistics, theoretical econometrics and bio-mathematics interested in mathematical objects occurring in the analysis of partially identified structures. The mathematical core of these ubiquitous structures has an impact on all three research areas and is expected to lead to the development of new algorithms for solving such problems.

Mathematics Subject Classification (2010): 62G08, 62G05, 68T05, 68Q32, 62P20, 60D05, 60E15.

Introduction by the Organisers

The workshop ”Mathematical Statistics of Partially Identified Objects” was organized by Victor Chernozhukov (Cambridge MA), Wolfgang Härdle (Berlin), Joel Horowitz (Evanston) and Ya’acov Ritov (Jerusalem) and was attended by 23 participants. The program included 21 talks of 60 minutes each, including discussions.

The workshop brought together mathematical statisticians, theoretical econometricians, and bio-mathematicians to understand and further develop the mathematical core of partially identified objects or structures. Partial identification is of ubiquitous nature in the analysis of structural models. Such analysis is relevant in many fields of applied mathematics. For example, when data are generated as outcomes of optimal discrete choices, moment inequalities do not identify the parameter, but they can be highly informative about it by restricting it to lie in the so called identification region. This phenomenon raises a variety of important
mathematical, statistical, and computational problems that were explored in the workshop.

Partial identification also arises in clinical trials. If some subjects do not take the treatments to which they are assigned or if there is attrition from the trial then the effects of treatment on the outcome variable are not identified unless one makes untestable assumptions about the attrition process. These identification problems are frequently explored through sensitivity analysis. A more thorough mathematical characterization of the described effects can be achieved by computing entire identification regions. Similar problems arise in survey research due to survey nonresponse. In the analysis of survey data, nonresponse is often dealt with by modeling the nonresponse process (e.g., assuming nonresponse is random conditional on observed covariates), but the models are not testable empirically and can lead to highly misleading conclusions.

Another example of partial identification is the prediction of the effects of a policy decision following, say, a clinical trial. Even if the results of a trial are not complicated by attrition or non-conformance to assigned treatments, they do not provide point predictions of what outcomes will occur in the general population. A clinical trial, at best, gives the average effect of treatment on a randomly selected group of individuals with the relevant disease or medical condition. However, once a drug or device is approved, those who receive are not randomly selected. Rather, they are chosen through a complicated and poorly understood process involving advice from medical professionals and the preferences of the patients. Consequently, the predicted effects of the new drug in the population are not identified.

The mathematical core problem is the characterization followed by computation of the identification regions. It is interesting to know whether computable identification regions can be formulated using modern tools from such seemingly unconnected fields of mathematics as the theory of the optimal mass transportation and the theory of random sets. These approaches have recently emerged as powerful tools in identification analysis, replacing earlier more primitive methods based on elementary algebraic manipulations. As a result, today, sharp and highly non-trivial bounds on parameters of games with multiple equilibria can be formulated using the random set theory and optimal transportation methods, substantially improving upon earlier non-sharp identification regions obtained for these models. There are also interesting relations to von Neumann’s method of alternating projections and maximum entropy methods.

Another challenging problem is estimating and performing inference on identification regions using available finite data. For example, the properties of the likelihood and other methods for performing inference on functionals of the parameter under partial identification are not yet well understood. Another example is the statistical theory of these methods and the appropriateness of convergence notions for stochastic programs, such as epi-convergence and related concepts, which have been developed in variational analysis and operations research. A number of interesting questions also arises in relation to the failure of conventional inferential
methods, e.g., bootstrap, due to limit distributions of relevant inferential statistics failing to be continuous with respect to the underlying probability measures. A number of methods have been suggested to remedy this failure; discussions around the theory of such methods with the vision of generating solutions that are both theoretically sound and practically relevant were at the center of the workshop.
## Workshop: Mathematical Statistics of Partially Identified Objects

### Table of Contents

Federico A. Bugni (joint with Ivan A. Canay, Xiaoxia Shi)

*Specification Tests for Partially Identified Models defined by Moment Inequalities* ................................................................. 1159

Peter Bühlmann

*Restricted structural equation models and improved bounds for high-dimensional causal inference* .............................. 1161

Ivan A. Canay (joint with Andrés Santos and Azeem M. Shaikh)

*On the Testability of Identification in Some Nonparametric Models with Endogeneity* ...................................................... 1162

Andrew Chesher and Adam M. Rosen

*Generalized Instrumental Variable Models* ............................... 1166

Holger Dette (joint with Stanislav Volgushev, Melanie Birke, Natalie Neumeyer)

*Significance testing in quantile regression* .................................. 1167

Alfred Galichon (joint with Victor Chernozhukov and Marc Henry)

*Higher dimensional quantiles and partial identification* .............. 1170

Maria Grith (joint with Wolfgang K. Härdle, Ya’acov Ritov)

*Remarks on the EPK Puzzle. The (Im)Possibility of Demixing* ....... 1173

Wolfgang Härdle (joint with Piotr Majer)

*Multidimensional statistical analysis of fMRI data in risk perception and investment decision study* ......................... 1176

Joel Horowitz (joint with Joachim Freyberger)

*Indentification and Shape Restrictions* ...................................... 1178

Tatiana Krivobokova

*Empirical Bayesian tuning parameters* ....................................... 1179

Sokbae Lee (joint with Kyungchul Song, Yoon-Jae Whang)

*Testing for a General Class of Functional Inequalities* .............. 1180

Yuan Liao (joint with Anna Simoni)

*Semi-parametric Bayesian Partially Identified Models* ............... 1181

Enno Mammen (joint with Ingrid van Keilegom, Kyusang Yu)

*Nonparametric Tests for Regression Quantiles* ......................... 1185
Nicolai Meinshausen (joint with Rajen Shah)  
Min-wise hashing for large-scale regression analysis. Computational limits of identifiability ........................................ 1190

Ilya Molchanov  
Selected topics on selections of random sets  ......................... 1190

Francesca Molinari (joint with Haim Bar)  
Computation of Sets Via Data Augmentation and Support Vector Machines ................................................................. 1192

Whitney Newey (joint with Jerry A. Hausman)  
Individual Heterogeneity and Average Welfare  ................. 1194

Vladimir Spokoiny (joint with Bill E. Xample, Max Muster)  
Identification and critical dimension in semiparametric estimation ..... 1195

Ngoc Mai Tran (joint with Maria Osipenko, Wolfgang Härdle)  
Principal Component Analysis in an Asymmetric Norm ............... 1197

Weining Wang (joint with Yan Fan, Wolfgang Karl Härdle, and Lixing Zhu)  
Composite Quantile Regression for the Single-Index Model ........... 1200