

# **Lecture Notes in Statistics**

**203**

Edited by P. Bickel, P.J. Diggle, S.E Fienberg, U. Gather,  
I. Olkin, S. Zeger



Pierre Alquier • Eric Gautier • Gilles Stoltz  
Editors

# Inverse Problems and High-Dimensional Estimation

Stats in the Château Summer School,  
August 31 - September 4, 2009

 Springer

*Editors*

Pierre Alquier  
Université Paris 7  
Laboratoire de Probabilités et  
Modèles Aléatoires  
175 rue du Chevaleret  
75205 Paris Cedex 13  
France  
[alquier@math.jussieu.fr](mailto:alquier@math.jussieu.fr)

Eric Gautier  
ENSAE  
Centre de Recherche en Economie et Statistique  
3 avenue Pierre Larousse  
92245 Malakoff  
France  
[eric.gautier@ensae-paristech.fr](mailto:eric.gautier@ensae-paristech.fr)

Gilles Stoltz  
HEC Paris  
Department of Economics and Decision Sciences  
1 rue de la Libération  
78351 Jouy-en-Josas  
&  
Ecole Normale Supérieure  
Département de Mathématiques et Applications  
45 rue d'Ulm  
75005 Paris  
France  
[stoltz@hec.fr](mailto:stoltz@hec.fr)

ISSN 0930-0325  
ISBN 978-3-642-19988-2      e-ISBN 978-3-642-19989-9  
DOI 10.1007/978-3-642-19989-9  
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011930794

© Springer-Verlag Berlin Heidelberg 2011

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

*Cover design:* eStudio Calamar S.L.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

## The “Stats in the Château” Summer School

The “Stats in the Château” summer school was held at the CRC château on the campus of HEC Paris, Jouy-en-Josas, France, from August 31 to September 4, 2009. It was organized jointly by faculty members of three French academic institutions: ENSAE ParisTech, Ecole Polytechnique ParisTech, and HEC Paris. These institutions cooperate through a scientific foundation devoted to the decision sciences.

The summer school brought together about 70 researchers and PhD students in economics, statistics, mathematics and computer science, all interested in both mathematical statistics and applications to economics. The motto was that economics is a source of interesting new problems for statisticians and that, conversely, recent statistical methods, sometimes motivated by other fields, can be used for quantitative analysis in economics. The goal was therefore to introduce the audience both to some modern sets of methods and to a wide range of their applications to economics, and to foster discussions between statisticians and economists.

The scientific content of the summer school focused on two themes: inverse problems and high-dimensional estimation. Two courses were given, one by Laurent Cavalier (Université Aix-Marseille I) on ill-posed inverse problems, and one by Victor Chernozhukov (Massachusetts Institute of Technology) on high-dimensional estimation with applications to economics. Ten invited lecturers – whose names can be found in the appendix of this volume together with the titles of their talk – illustrated the two courses and provided either reviews of the state of the art in the field or of applications and original research contributions. The opportunity was also given to participants to present their own findings either in the form of a contributed talk or during a poster session held on the first day of the summer school.

## Inverse Problems and High-Dimensional Estimation

The first theme of the summer school was ill-posed statistical inverse problems. This is already considered an important problem in many areas of science, and also became an important issue in econometrics about ten years ago. Nonparametric specifications are important to allow for flexible models. Statistical inverse problems are problems from nonparametric statistics. A wide class of models from economics can be formulated as inverse problems, that is, as a relation between a function, directly related to the observations, and a structural functional parameter. Examples include the estimation of the distributions of the following objects: types in a game-theoretical model where only actions of the players are observed; random coefficients accounting for unobserved heterogeneity; the pricing operator given observed option prices; a regression function in the presence of endogenous regressors; etc. Because inversion often leads to a lack of continuity, the inference requires some suitable regularization. Theoretical properties such as optimal rates of convergence and adaptation are important to study.

The second theme was high-dimensional estimation. High-dimensionality corresponds to the case where the parameter of interest has a dimension  $p$  possibly much larger than the sample size  $n$ . A lot of attention has been given to this setting in recent years in the statistics and machine learning communities. In this setting, parsimonious models can still be estimated. Parsimony is also referred to as sparsity and corresponds to the case where, though the number of parameters is very large, only a small number of them are non-zero. Results can often be extended to the case where most parameters are too small to matter. This is a setting often encountered in the social sciences. Several techniques have been developed to extract relevant parameters from large vectors, in particular, the Lasso, the Dantzig selector, and Bayesian-type methods. These techniques have been widely implemented in imaging and bioinformatics. At the time of the summer school the use of the above-mentioned methods in economics, while widely applicable, was very limited.

## The Proceedings

After the summer school, the twelve researchers giving the lectures or the invited talks were given the opportunity to contribute to the present volume. The aim was to provide an accessible but rigorous mathematical introduction to these two modern sets of problems from statistics and econometrics, and to present applications to quantitative problems in economics. The intended audience is the same as that of the summer school: young researchers, e.g., PhD students in statistics and economics, or more senior researchers from related fields.

The book brings together contributions from five invited speakers with their coauthors, among them, the two lecturers. Laurent Cavalier provides detailed lecture notes on ill-posed statistical inverse problems while Victor Chernozhukov reviews Lasso-based methods for estimating high-dimensional regression models with ap-

plications to empirical economic problems. These lecture notes are illustrated and further developed by three other contributions. On the one hand, Jean-Pierre Florens discusses the case of nonparametric estimation with endogenous variables using instrumental variables. On the other hand, Felix Abramovich and Ya'acov Ritov respectively present a model selection and a Bayesian viewpoint on high-dimensional estimation.

Interested readers can find the slides of most of the invited and contributed talks, as well as the videotape of the first lecture by Laurent Cavalier, on the website of the summer school <http://www.hec.fr/statsinthechateau>.

## Acknowledgments

The scientific committee of the summer school included

Christian Gourieroux	(ENSAE-CREST – Université de Toronto)
Yuichi Kitamura	(Yale University)
Alexandre Tsybakov	(ENSAE-CREST)

The summer school was mainly funded by the “Groupement d’Intérêt Scientifique: Sciences de la Décision”, a scientific foundation devoted to the decision sciences, gathering Ecole Polytechnique ParisTech, ENSAE ParisTech and HEC Paris.



It was also supported by the EDF-Calyon “Finance et développement durable” (finance and sustainable development) chair.

The faculty members in charge of the local organization were

Pierre Alquier	(ENSAE-CREST – Université Paris Diderot)
Veronika Czellar	(HEC Paris)
Alfred Galichon	(Ecole Polytechnique ParisTech)
Eric Gautier	(ENSAE-CREST)
Gilles Stoltz	(CNRS – Ecole normale supérieure – HEC Paris)

and they were assisted by

Nathalie Beauchamp	(HEC Paris)
Claudine Tantillo	(HEC Paris)

Paris, January 2011

Pierre Alquier  
Eric Gautier  
Gilles Stoltz



# Contents

## Part I Lecture Notes on Inverse Problems

<b>1</b>	<b>Inverse Problems in Statistics</b> .....	<b>3</b>
	Laurent Cavalier	
1.1	Inverse Problems .....	4
1.1.1	Introduction .....	4
1.1.2	Statistical Inverse Problems .....	6
1.1.3	Linear Inverse Problems with Random Noise .....	7
1.1.4	Basic Notions on Operator Theory .....	9
1.1.5	Singular Value Decomposition and Sequence Space Model .....	13
1.1.6	Examples .....	16
1.1.7	Spectral Theory .....	23
1.2	Nonparametric Estimation .....	27
1.2.1	Minimax Approach .....	28
1.2.2	Regularization Methods .....	29
1.2.3	Classes of Functions .....	38
1.2.4	Rates of Convergence .....	42
1.2.5	Comparison Between Deterministic and Stochastic Noise	48
1.3	Adaptation and Oracle Inequalities .....	53
1.3.1	Minimax Adaptive Procedures .....	53
1.3.2	Oracle Inequalities .....	55
1.3.3	Model Selection .....	57
1.3.4	Universal Optimality .....	75
1.4	Conclusion .....	84
1.4.1	Summary .....	84
1.4.2	Discussion .....	86
1.4.3	Open Problems .....	87
	References .....	92

## Part II Invited Contribution on Inverse Problems

<b>2</b>	<b>Non-parametric Models with Instrumental Variables</b> . . . . .	99
	Jean-Pierre Florens	
2.1	Introduction . . . . .	99
2.2	The Linear Model: Vectorial or Functional Data . . . . .	102
2.3	The Additively Separable Model and Its Extensions . . . . .	106
2.4	The Non-separable Models . . . . .	111
2.5	Some Extensions to Dynamic Models . . . . .	113
2.6	Conclusion . . . . .	115
	References . . . . .	116

## Part III Lecture Notes on High-Dimensional Estimation

<b>3</b>	<b>High Dimensional Sparse Econometric Models: An Introduction</b> . . . .	121
	Alexandre Belloni and Victor Chernozhukov	
3.1	The High Dimensional Sparse Econometric Model . . . . .	121
3.2	The Setting and Estimators . . . . .	127
	3.2.1 The Model . . . . .	127
	3.2.2 LASSO and Post-LASSO Estimators . . . . .	129
	3.2.3 Intuition and Geometry of LASSO and Post-LASSO . . . .	131
	3.2.4 Primitive Conditions . . . . .	133
3.3	Analysis of LASSO . . . . .	135
3.4	Model Selection Properties and Sparsity of LASSO . . . . .	137
	3.4.1 Sparsity Bounds . . . . .	137
	3.4.2 Perfect Model Selection Results . . . . .	138
3.5	Analysis of Post-LASSO . . . . .	139
3.6	Estimation of Noise Level . . . . .	141
3.7	Monte Carlo Experiments . . . . .	142
3.8	Application to Cross-Country Growth Regression . . . . .	146
	Appendix . . . . .	147
3.9	Proofs . . . . .	147
3.10	Auxiliary Lemmas . . . . .	153
	References . . . . .	155

## Part IV Invited Contributions on High-Dimensional Estimation

<b>4</b>	<b>Model Selection in Gaussian Regression for High-Dimensional Data</b> .	159
	Felix Abramovich and Vadim Grinshtein	
4.1	Introduction . . . . .	159
4.2	Preamble . . . . .	161
4.3	Nearly-Orthogonal Design . . . . .	163
4.4	Multicollinear Design . . . . .	165
4.5	Bayesian Perspective . . . . .	166
4.6	Concluding Remarks . . . . .	168
	References . . . . .	169

**5 Bayesian Perspectives on Sparse Empirical Bayes Analysis (SEBA) . . . 171**  
 Natalia Bochkina and Ya'acov Ritov

- 5.1 Introduction . . . . . 171
- 5.2 The Lassoes Procedure . . . . . 174
  - 5.2.1 Persistency and Bayesian Interpretation . . . . . 174
  - 5.2.2 Restricted Eigenvalues Condition and Oracle Inequalities 177
- 5.3 Group Lasso: Bayesian Perspective . . . . . 180
- 5.4 RING Lasso: Bayesian Perspective . . . . . 182
- References . . . . . 189

**Part V Invited and Contributed Talks Given During the Summer School**

- A List of the Courses . . . . . 193**
- B List of the Invited Talks . . . . . 195**
- C List of the Contributed Talks and Posters . . . . . 197**



# List of Contributors

Felix Abramovich

Tel Aviv University, Department of Statistics & Operations Research, Ramat Aviv,  
Tel Aviv 69978, Israel, e-mail: [felix@post.tau.ac.il](mailto:felix@post.tau.ac.il)

Alexandre Belloni

Duke University, Fuqua School of Business, 100 Fuqua Drive, Durham, NC 27708,  
USA, e-mail: [abn5@duke.edu](mailto:abn5@duke.edu)

Natalia Bochkina

University of Edinburgh, School of Mathematics, King's Buildings, Mayfield Road,  
Edinburgh, EH9 3JZ, UK, e-mail: [N.Bochkina@ed.ac.uk](mailto:N.Bochkina@ed.ac.uk)

Laurent Cavalier

Université Aix-Marseille 1, LATP, CMI, 39 rue Joliot-Curie, 13453 Marseille,  
France, e-mail: [cavalier@cmi.univ-mrs.fr](mailto:cavalier@cmi.univ-mrs.fr)

Victor Chernozhukov

Massachusetts Institute of Technology, Department of Economics, 50 Memorial  
Drive, Cambridge, MA 02142, USA, e-mail: [vchern@mit.edu](mailto:vchern@mit.edu)

Jean-Pierre Florens

Université Toulouse 1 & Toulouse School of Economics, GREMAQ & IDEI, 21  
allée de Brienne, 31000 Toulouse, France, e-mail: [florens@cict.fr](mailto:florens@cict.fr)

Vadim Grinshtein

The Open University of Israel, Department of Mathematics, P.O.Box 808, Raanana  
43107, Israel, e-mail: [vading@openu.ac.il](mailto:vading@openu.ac.il)

Ya'acov Ritov

The Hebrew University of Jerusalem, Department of Statistics, Mount Scopus,  
Jerusalem 91905, Israel, e-mail: [yaacov.ritov@gmail.com](mailto:yaacov.ritov@gmail.com)

