



Universiteit  
Leiden  
The Netherlands

## Prediction accuracy and stability of regression with optimal scaling transformations

Kooij, A.J. van der

### Citation

Kooij, A. J. van der. (2007, June 27). *Prediction accuracy and stability of regression with optimal scaling transformations*. Leiden. Retrieved from <https://hdl.handle.net/1887/12096>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/12096>

**Note:** To cite this publication please use the final published version (if applicable).

Prediction Accuracy and Stability  
of Regression with  
Optimal Scaling Transformations

Van der Kooij, Anita Jolande,  
Prediction Accuracy and Stability of Regression with Optimal Scaling  
Transformations.  
Dissertation Leiden University — With ref. — With Summary in Dutch.  
Subject headings: nonlinear regression; CATREG; optimal scaling;  
transformations; local minima; prediction accuracy; regularization;  
.632 Bootstrap; Ridge regression; Lasso; Elastic Net

ISBN 978-90-9021936-3

© 2007 Anita J. van der Kooij

Printed by Mostert en van Onderen, Leiden

Prediction Accuracy and Stability of  
Regression with Optimal Scaling  
Transformations

Proefschrift

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,

op gezag van de Rector Magnificus prof.mr. P.F. van der Heijden,

volgens besluit van het College voor Promoties

te verdedigen op woensdag 27 juni 2007

klokke 16.15 uur

door

Anita J. van der Kooij

geboren te Boskoop

in 1961

## PROMOTIECOMMISSIE

Promotor            Prof. dr. J.J. Meulman

Referent            Prof. J.H. Friedman, Ph.D., Stanford University, USA

Overige Leden      Prof. dr. W.J. Heiser  
                         Prof. dr. M.H. van IJzendoorn  
                         Dr. ing. P.H.C. Eilers  
                         Prof. dr. R.D. Gill

# Contents

|          |                                                                                                               |           |
|----------|---------------------------------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Introduction</b>                                                                                           | <b>1</b>  |
| 1.1      | Regression methods for nonlinearly related data . . . . .                                                     | 2         |
| 1.2      | The CATREG method . . . . .                                                                                   | 3         |
| 1.2.1    | Optimal scaling levels . . . . .                                                                              | 5         |
| 1.2.2    | Estimation of transformations . . . . .                                                                       | 8         |
| 1.2.3    | Local minima . . . . .                                                                                        | 10        |
| 1.2.4    | Prediction accuracy, stability, and regularization . . . .                                                    | 11        |
| 1.3      | Outline . . . . .                                                                                             | 12        |
| <b>2</b> | <b>Local Minima in Regression with Optimal Scaling Transformations</b>                                        | <b>15</b> |
| 2.1      | Introduction . . . . .                                                                                        | 16        |
| 2.2      | Model . . . . .                                                                                               | 17        |
| 2.3      | Algorithm . . . . .                                                                                           | 18        |
| 2.4      | Local minima . . . . .                                                                                        | 21        |
| 2.4.1    | A probable cause of local minima . . . . .                                                                    | 21        |
| 2.4.2    | Strategies to obtain the global minimum . . . . .                                                             | 23        |
| 2.5      | Simulation study . . . . .                                                                                    | 26        |
| 2.5.1    | Design . . . . .                                                                                              | 26        |
| 2.5.2    | Results . . . . .                                                                                             | 28        |
| 2.6      | Conclusion . . . . .                                                                                          | 35        |
| <b>3</b> | <b>Prediction Accuracy of Regression with Optimal Scaling transformations: The .632 Bootstrap with CATREG</b> | <b>37</b> |
| 3.1      | Introduction . . . . .                                                                                        | 37        |
| 3.2      | CATREG: Regression with Optimal Scaling Transformations .                                                     | 40        |
| 3.2.1    | CATREG model . . . . .                                                                                        | 40        |
| 3.2.2    | CATREG algorithm . . . . .                                                                                    | 41        |
| 3.3      | Estimation of expected prediction error . . . . .                                                             | 44        |

|          |                                                                                                                                             |           |
|----------|---------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| 3.3.1    | The .632 bootstrap with linear regression . . . . .                                                                                         | 45        |
| 3.3.2    | The .632 bootstrap with CATREG . . . . .                                                                                                    | 46        |
| 3.4      | Performance of CATREG and six other regression with transformation methods: Prediction accuracy in the analysis of the Ozone data . . . . . | 47        |
| 3.4.1    | Comparison between CATREG and Box-Tidwell Method, Full Additive Model, Stepwise Full Additive Model, TURBO, and BRUTO . . . . .             | 48        |
| 3.4.2    | Comparison between CATREG and ACE . . . . .                                                                                                 | 50        |
| 3.5      | Prediction accuracy for different scaling levels in CATREG . .                                                                              | 53        |
| 3.6      | The effect of the number of observations on the expected prediction error . . . . .                                                         | 58        |
| 3.7      | Conclusions . . . . .                                                                                                                       | 61        |
| 3.8      | Computational note . . . . .                                                                                                                | 63        |
| <b>4</b> | <b>Regularization with Ridge penalties, the Lasso, and the Elastic Net for Regression with Optimal Scaling Transformations</b>              | <b>65</b> |
| 4.1      | Introduction . . . . .                                                                                                                      | 66        |
| 4.2      | Ridge penalties, the Lasso, and the Elastic Net for linear regression . . . . .                                                             | 67        |
| 4.3      | Ridge penalties, the Lasso, and the Elastic Net with CATREG                                                                                 | 70        |
| 4.3.1    | Updating the regularization regression coefficients in CATREG . . . . .                                                                     | 71        |
| 4.3.2    | Paths for the coefficients . . . . .                                                                                                        | 72        |
| 4.3.3    | Finding the transition points in CATREG . . . . .                                                                                           | 74        |
| 4.3.4    | Including nonlinear transformations in the CATREG algorithm . . . . .                                                                       | 75        |
| 4.4      | Selection of the optimal penalty parameter . . . . .                                                                                        | 75        |
| 4.4.1    | Illustration . . . . .                                                                                                                      | 76        |
| 4.5      | Shrinking a nominal variable versus shrinking it's categories . .                                                                           | 80        |
| 4.6      | Discussion . . . . .                                                                                                                        | 86        |
| <b>5</b> | <b>Application of the CATREG-Lasso: Severity of Bulimia Nervosa - Components of the Syndrome and Definition of Therapy Outcome</b>          | <b>91</b> |
| 5.1      | Introduction . . . . .                                                                                                                      | 92        |
| 5.2      | Method . . . . .                                                                                                                            | 94        |
| 5.2.1    | Sample and instruments . . . . .                                                                                                            | 94        |
| 5.2.2    | Self and Expert Ratings . . . . .                                                                                                           | 97        |
| 5.2.3    | Statistical Analysis . . . . .                                                                                                              | 97        |

|          |                                                                       |            |
|----------|-----------------------------------------------------------------------|------------|
| 5.3      | Results . . . . .                                                     | 101        |
| 5.3.1    | Model selection . . . . .                                             | 101        |
| 5.3.2    | Discrimination between health and pathology . . . . .                 | 104        |
| 5.4      | Discussion . . . . .                                                  | 105        |
| <b>6</b> | <b>General Discussion</b>                                             | <b>109</b> |
| 6.1      | A short retrospect . . . . .                                          | 109        |
| 6.2      | Topics for further research . . . . .                                 | 114        |
| 6.2.1    | Transformations towards independence and regularization               | 114        |
| 6.2.2    | The effect of shrinking on low-frequency categories . . .             | 114        |
| 6.2.3    | Regularization in Discriminant Analysis . . . . .                     | 115        |
| 6.2.4    | CATREG with cluster restrictions . . . . .                            | 115        |
|          | <b>Appendix A CATREG Algorithm</b>                                    | <b>117</b> |
|          | <b>Appendix B CATREG sections from SPSS Categories® 11.0</b>          | <b>131</b> |
|          | <b>Appendix C Notation</b>                                            | <b>179</b> |
|          | <b>References</b>                                                     | <b>181</b> |
|          | <b>Summary in Dutch (Samenvatting)</b>                                | <b>189</b> |
|          | <b>Curriculum vitae</b>                                               | <b>197</b> |
|          | <b>Overview of Applications</b>                                       |            |
|          | CATREG for Diamonds data . . . . .                                    | 6          |
|          | Prediction accuracy for Ozone data . . . . .                          | 47         |
|          | Effect of number of observations on prediction accuracy for           |            |
|          | Demographic data . . . . .                                            | 58         |
|          | Linear Lasso for Diabetes data . . . . .                              | 73         |
|          | Linear and nonlinear Ridge, Lasso, and Elastic Net for                |            |
|          | Prostate cancer data . . . . .                                        | 76         |
|          | Nonlinear Lasso and dummies-Lasso for Breast cancer data . . . . .    | 82         |
|          | Nonlinear Lasso and .632 bootstrap for Bulimia Nervosa data . . . . . | 91         |

