IWSM 2011

Proceedings of the

26th International Workshop on Statistical Modelling

Valencia (Spain), July 11-15, 2011

Editors:
David Conesa
Anabel Forte
Antonio López-Quílez
Facundo Muñoz

Proceedings of the 26th International Workshop on Statistical Modelling.
València, July 11-15, 2011
David Conesa, Anabel Forte, Antonio López-Quílez, Facundo Muñoz, eds.
València 2011.
ISBN 978-84-694-5129-8
Depósito Legal V-2498-2011

Editors:

David Conesa[1], David.V.Conesa@uv.es
Anabel Forte[2], forte@eco.uji.es
Antonio López-Quílez[1], Antonio.Lopez@uv.es
Facundo Muñoz[1], Facundo.Munoz@uv.es

[1] Departament d'Estadística i Investigació Operativa
Universitat de València (Estudi General)
Facultat de Matemàtiques
Dr. Moliner 50, 46100 Burjassot, Spain.

[2] Departamento de Economía
Universitat Jaume I
Facultad de Ciencias Jurídicas y Económicas
Campus del Riu Sec, E-12071 Castelló de la Plana, Spain.

Cover photo: Victor Roda

Printed by Copiformes S.L.

Contents

Part 1. Invited papers

Part 2. Contributed papers

impacts upon subject classification more than the model's fixed effects, as these are simply the average of individual fitted curves whilst subject classification is based on individual curves. Subject classification is key to the utility of GMMs; models that capture model-generated autocorrelation within the GMM framework are thus preferred. Whilst the exact choice of parameterization remains open, our findings suggest that some kind of explicit modelling of autocorrelation is warranted in these types of models. In any event, correct parameterization of the random structure is needed for growth mixture modelling of outcomes that exhibit less within-subject than between-subject heterogeneity.

Berkey, C.S. & Colditz, G.A. (2007). Adiposity in adolescents: change in actual BMI works better than change in BMI z score for longitudinal studies. *Ann. Epidemiol.*, **17**(1), 44-50.

Bollen, K. & Curran, P. (2006). *Latent curve models*, 2nd ed. New York, Wiley.

Duncan, T.E., Duncan, S.E., & Stryker, L.A. (2006). *An introduction to latent variable growth curve modeling*, 2nd ed. Mahwah, NJ, Laurence Erlbaum Associates Inc.

Goodman, E., Adler, N.E., Daniels, S.R., Morrison, J.A., Slap, G.B., & Dolan, L.M. (2003). Impact of objective and subjective social status on obesity in a biracial cohort of adolescents. *Obes. Res.*, **11**(8), 1018-1026.

Kreuter, F. & Muthen, B. (2008). Analyzing Criminal Trajectory Profiles: Bridging Multilevel and Group-based Approaches Using Growth Mixture Modeling. *Journal of Quantitative Criminology*, **24**(1), 1-31.

Kuczmarski, R.J., Ogden, C.L., Grummer-Strawn, L.M., Flegal, K.M., Guo, S.S., Wei, R., Mei, Z., Curtin, L.R., Roche, A.F., & Johnson, C.L. (2000). CDC growth charts: United States. *Adv. Data.* **314**, 1-27.

Li, C., Goran, M.I., Kaur, H., Nollen, N., & Ahluwalia, J.S. (2007). Developmental trajectories of overweight during childhood: role of early life factors. *Obesity. (Silver. Spring)*, **15**(3), 760-771.

Mustillo, S., Worthman, C., Erkanli, A., Keeler, G., Angold, A., & Costello, E.J. (2003). Obesity and psychiatric disorder: developmental trajectories. *Pediatrics*, **111**(4-1), 851-859.

Needham, B.L., Epel, E.S., Adler, N.E., & Kiefe, C. (2010). Trajectories of change in obesity and symptoms of depression: the CARDIA study. *Am. J. Public Health*, **100**(6), 1040-1046.

Modeling swimming marks through Blocks and POT methods

Dulce Gomes[1], Júlia Teles[2], Luísa Canto e Castro[3]

[1] Departamento de Matemática/CIMA, Escola de Ciências e Tecnologia, Universidade de Évora, Rua Romão Ramalho, 59, 7000-671 Évora, Portugal (email: dmog@uevora.pt)

[2] Departamento de Métodos Matemáticos/CIPER, Faculdade de Motricidade Humana, Universidade Técnica de Lisboa, Estrada da Costa, 1495-688 Cruz Quebrada-Dafundo, Portugal (email: jteles@fmh.utl.pt)

[3] Departamento de Estatística e Investigação Operacional/CEAUL, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Edifício C6, 1749-016 Lisboa, Portugal (email: luisa.loura@deio.fc.pt)

Abstract: The swimming marks in the 100m men's freestyle long course are modelled using extreme value theory. Using the statistical package R, extreme value and generalized Pareto models were adjusted in order to estimate the left endpoint of these models. The left endpoint can be interpreted as the best mark that can ever be reached, admitting that swimming pool conditions, athlete's equipment and training methods remain the same.

Keywords: Extreme value models; Generalized Pareto models; Blocks method; Peaks over threshold method; Swimming marks.

1 Introduction

Extreme value models are frequently used for the analysis of samples of maximum or minimum and generalized Pareto models are commonly used to analysing the samples of exceedances over a high or low threshold.
The swimming marks of 100m men's freestyle (long course), that appear in FINA ("La Federation Internationale de Natation") Website (www.fina.org), are the personal best in a very large sample of marks, so we could say that we are in the presence of extreme value — in this case minima. In this sense, we consider analysing and modeling this type of data by means of extreme value models.
Using two methods of extreme value analysis — the blocks method (De Haan and Ferreira, 2001) and the peaks over threshold (POT) method (Pickands, 1975; Robinson and Tawn, 1995) — we are going to model the swimming marks and estimate their left endpoint. This left endpoint can be interpreted as the best mark that can ever be reached, admitting that swimming pool conditions, athlete's equipment and methods of training remain the same.

TABLE 1. The three best annual marks (in seconds), through 1948 to 2010, of Men's Long Course World Records in 100m freestyle.

year	rank	mark	athlete	nationality
1948	1	57.3	Wally Ris	USA
1948	2	57.6	Keith Carter	USA
1948	3	57.8	Alan Ford	USA
...	...	...	...	...
1999	1	48.35	Pieter van den Hoogenband	NED
1999	2	48.73	Michael Klim	AUS
1999	3	48.82	Alexander Popov	RUS
2000	1	47.84	Pieter van den Hoogenband	NED
2000	2	48.18	Michael Klim	AUS
2000	3	48.27	Alexander Popov	RUS
...	...	...	...	...
2010	1	48.54	Simon Burnett	GBR
2010	2	48.56	William Meynard	FRA
2010	3	48.69	Kyle Richardson	AUS

A draft of the dataset with the three best annual marks (in seconds) of the men's long course world records in 100m freestyle are presented in the Table 1. The information was available from 1948 to 2010, with missing value for 1950 and 1951.
In Figure 1, the marks are plotted against the year and by ranking. As we expected, there is a decreasing trend in the marks. So, the relevant question is: Until when these marks could fall? In order to give answer to this question two different approaches of extreme value theory were used.

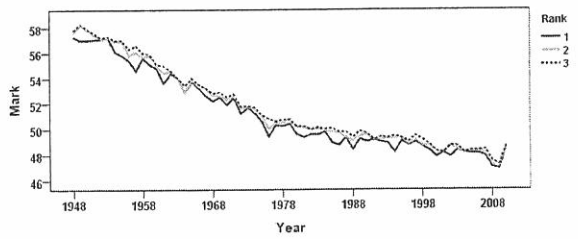

FIGURE 1. Scatter plot of mark against year by ranking.

2 Two different approaches

To apply the extreme value methodology we must have independent and identically distributed observations (Reiss and Thomas, 2001). As we can observe in Table 1, there are some athletes (e.g., Michael Klim and Alexander Popov) which contributed with more than one mark. So to use this type

of analysis we only select the best mark of each athlete. In order to adjust an extreme value or generalized Pareto model, the trend also needs to be removed.
In the POT approach the inference is based in the exceedances over a high threshold that is unknown. Our empirical way of choosing this threshold was through the analysis of the diagram of the shape parameter's estimates. To apply the POT method we adjust a model in the family of generalized Pareto models. For different shape, location and scale parameters we obtain three different submodel families: exponential, Pareto and beta.
In the block method we adjust a model in the family of extreme value models. Also depending on the shape, location and scale parameters we could reach the Gumbel, Fréchet or Weibull submodels.

Acknowledgments: This research was partially support by the Center of Mathematics and Applications, University of Évora, by the Interdisciplinary Centre for the Study of Human Performance, Technical University of Lisbon, and by the Center of Statistics and Applications, University of Lisbon, through the Programs FCT/POCTI, FCT/POCI2010 and POCI/FEDER.

References

De Haan, L., and Ferreira, A. (2006). Extreme Value Theory: An Introduction. Boston: Springer.

Pickands, J. (1995). Statistical inference using extreme value order statistics. *Annals of Statistics*, **3**, 119-131.

R Development Core Team (2010). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria. (Available at http://www.R-project.org/).

Reiss, R.D., and Thomas, M. (2001). *Statistical Analysis of Extreme Values*, 2nd edition. Basel: Birkhäuser.

Robinson, M.E., and Tawn, J.A. (1995). Statistics for expected athletics records. *Applied Statistics*, **44**, 499-511.