Asymmetric least squares smoothing and life expectancy frontiers

Sabine K. Schnabel[1] and Paul H.C. Eilers[2]

[1] Max Planck Institute for Demographic Research, Rostock, Germany
[2] Utrecht University, Faculty of Social and Behavioural Sciences, Utrecht, The Netherlands

In the light of worldwide increasing life expectancy, demographers are interested in analyzing its limits. When studying a scatterplot of observed (time series) data such as the development of life expectancy over time and as a function of socio-economic indicators, we might not only be interested in the trend, but also in the spread around it. Quantile smoothing (QS) is an effective and popular tool for this purpose. QS is based on asymmetrically weighting the sum of absolute values of residuals. We propose asymmetrically weighting the sum of squares of residuals that leads to so called expectiles as introduced in [3]. It is extremely easy to fit any asymmetric least squares (ALS) model: simply iterate between weighted regression and re-compute the weights. The goal function is convex, so a unique minimum is guaranteed. We combine ALS with P-splines as presented in [1].

As in any smoothing problem, it is important to have an automatic method to determine a good value of the smoothing parameter - in our case the weight of the penalty of P-splines. We present two methods: one is based on asymmetric cross-validation, the other adapts Schall's EM algorithm for estimating variance components. In the asymmetric variant of the classical cross-validation score we introduce a weight vector into the calculation and perform a grid search over the smoothing parameter λ in order to find the minimal score. Alternatively, we use the formal equivalence between penalized least squares smoothing and mixed models (see [4, 2]) and adapt the Schall algorithm from [5] that was originally designed for generalized linear mixed models. With this algorithm we estimate the variance components, i.e. variance of the errors and variance of the contrasts of the differentiated coefficient vector. The quotient of the two variances leads to the optimal smoothing parameter λ in an iterative process involving smoothing asymmetrically with λ and estimating variances (and a new λ) until convergence.

The proposed model gives attractive results for the frontier of life expectancy as a function of Gross Domestic Product per capita and time. Our contribution includes this application as an example. Furthermore the method was tested in an extensive simulation study.

Due to its rather simple concept and easy implementation, as a modification of classical ordinary least squares, our approach can be extended to multi-dimensional contexts and combined with mixed models.

References

[1] Eilers P.H.C., Marx B.D. (1996) Flexible smoothing with B-splines and penalties. Statistical Sciences 11:89-121.

[2] Lee Y., Nelder J.A., Pawitan Y. (2006) *Generalized linear models with random effects – unified analysis via H-likelihood*. London: Chapman & Hall

[3] Newey W.K., Powell J.L. (1987) Asymmetric least squares estimation and testing. Econometrica 55:819-847

[4] Pawitan Y. (2001) *In all likelihood.* Oxford: Oxford University Press

[5] Schall R. (1991) Estimation in Generalized Linear Models with random effects. Biometrika 78:719-727