On information criteria for estimating the number of knots in splines with fractional residuals

Beran, Jan
University of Konstanz, Department of Mathematics and Statistics
Universitätsstraße 10
78457 Konstanz, Germany
E-mail: jan.beran@uni-konstanz.de

Weiershäuser, Arno
University of Konstanz, Department of Mathematics and Statistics
Universitätsstraße 10
78457 Konstanz, Germany
E-mail: arno.weiershaeuser@uni-konstanz.de

Introduction

We consider the model

$$x(t) = \mu\left(\tfrac{t}{n}\right) + \xi_t \quad (t = 1, \ldots, n)$$

with μ denoting a piecewise polynomial function and $(\xi_t)_{t\in\mathbf{Z}}$ a stationary stochastic process. Thus

$$\mu(s) = \sum_{k=0}^{l}\sum_{j=1}^{p_k} a_{j,k}(s-\kappa_k)_+^{b_{j,k}}$$

with some knots $0 = \kappa_0 < \kappa_1 < \ldots < \kappa_l < 1$, regression coefficients $a_{j,k} \in \mathbf{R}$ and integer exponents $b_{j,k}$. The exponents $b_{j,k}$ are assumed to be positive for all $k > 0$ so that μ is continuous. The error process is given by

$$\xi_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$$

with $(c_j)_{j\in\mathbf{Z}} \in l^2(\mathbf{Z})$ and $(\varepsilon_t)_{t\in\mathbf{Z}}$ denoting an iid sequence such that $\mathbf{E}[\varepsilon_t] = 0$ and $\mathbf{E}[|\varepsilon_t|^r] < \infty$ for some $r > 2$. Define $\gamma(k) = \mathrm{Cov}(\xi_t, \xi_{t+k})$ and $D^2(n) = \mathrm{Var}\left(\sum_{t=1}^{n}\xi_t\right)$. The linear process ξ_t is assumed to satisfy one of the following conditions:

1. Short memory:

   $$\sum_{k\in\mathbf{Z}} |\gamma(k)| < \infty,\ 0 < \sum_{k\in\mathbf{Z}} \gamma(k) < \infty.$$

   In this case, we have

   $$D^2(n) \sim n \sum_{k\in\mathbf{Z}} \gamma(k). \tag{1}$$

2. long memory: there exists $N \in \mathbf{N}$, a slowly varying function L and $\alpha \in (0, 1)$ such that

   $$\gamma(k) = L(k)\, k^{-\alpha}\ (k \geq N).$$

   In this case, we have

   $$\sum_{k\in\mathbf{Z}} |\gamma(k)| = \infty$$

and

$$D^2(n) \sim \frac{2n^2\gamma(n)}{(1-\alpha)(2-\alpha)}. \tag{2}$$

3. antipersistence: there exists $N \in \mathbf{N}$, a slowly varying function L and some $\alpha \in (1,2)$ such that

$$\gamma(k) = -L(k)\,k^{-\alpha}\ (k \geq N),\ \sum_{k\in\mathbf{Z}} \gamma(k) = 0.$$

In this case, we have

$$D^2(n) \sim \frac{2n^2\,|\gamma(n)|}{(\alpha-1)(2-\alpha)}. \tag{3}$$

Furthermore, we assume that $\mathbf{E}\left(\xi_t^4\right) < \infty$ and $\mathrm{Var}\left(n^{-1}\sum_{t=1}^n \xi_t^2\right) \to 0$. In the case of antipersistence, we assume in addition that $r(2-\alpha) > 2$.

If the number of knots is known, both the regression coefficients and the position of the knots can be estimated consistently by least squares regression. The asymptotic distribution of the parameter estimates can be derived by non-linear least square techniques (see Gallant 1974, Feder 1975, Liu et al 1997 and Kim and Kim 2008 for the case of iid errors, Beran and Weiershäuser 2010 and Beran et al. 2011 for the cases of long memory and antipersistence).

In practice, the number of knots is typically unknown and therefore needs to be estimated. Yao (1988) showed consistency of the Schwarz criterion for piecewise constant spline functions and iid normal random variables. Liu et al. (1997) show that for locally exponentially bounded iid errors the number of knots can be estimated consistently by minimising a modified Schwarz criterion

$$MIC(l) = \log\left[S\left(\hat{\kappa}_1,\ldots,\hat{\kappa}_l\right)/(n-p^*)\right] + p^*\frac{c_0\left(\log(n)\right)^{2+\delta_0}}{n}.$$

Here, $S\left(\hat{\kappa}_1,\ldots,\hat{\kappa}_l\right)$ denotes the residuals sum of squares obtained under a model with l knots and p^* parameters in total, c_0 and δ_0 denote strictly positive constants.

Asymptotic results

The following definitions are used.

Definition 1 *The space of ordered l-tuples is denoted by*

$$\mathbf{S}^l = \left\{(k_1,\ldots,k_l) \in (0,1)^l : k_1 < k_2 < \ldots < k_l\right\}.$$

Moreover, define for all $\Delta > 0$

$$\mathbf{S}^l_\Delta = \left\{(k_1,\ldots,k_l) \in \mathbf{S}^l : |k_i - k_j| \geq \Delta\right\}.$$

Definition 2 *Let $l \in \mathbf{N}$, $p_k \in \mathbf{N}\setminus\{0\}$ and $b_{j,k} \in \mathbf{N}$ $(j = 1,\ldots,p_k,\ k = 0,\ldots,l)$. For each tripel $(l,(p_k),(b_{j,k}))$, define the parametric family $\mathbf{M} = \mathbf{M}_{l,(p_k),(b_{j,k})}$ of spline functions by*

$$\mathbf{M} = \left\{\mu : [0,1] \to \mathbf{R}, \mu(t) = \sum_{k=0}^{l}\sum_{j=1}^{p_k} a_{j,k}(t-\kappa_k)_+^{b_{j,k}}, \kappa_0 = 0, (\kappa_k)_{k=1}^l \in \mathbf{S}^l, a_{j,k} \in \mathbf{R}\right\}.$$

Likewise,

$$\mathbf{M}(\Delta) = \left\{\mu : [0,1] \to \mathbf{R}, \mu(t) = \sum_{k=0}^{l}\sum_{j=1}^{p_k} a_{j,k}(t-\kappa_k)_+^{b_{j,k}}, \kappa_0 = 0, (\kappa_k)_{k=1}^l \in \mathbf{S}^l_\Delta, a_{j,k} \in \mathbf{R}\right\}.$$

We refer to l as the number of knots *and to $p_{\mathbf{M}} = l + \sum_{k=0}^{l} p_k$ as the* total number of parameters *of* $\mathbf{M}$.

Definition 3 *Let* $\Lambda = \{\mathbf{M}_1, \ldots, \mathbf{M}_q\}$ *be a set of parametric spline families. We say that* Λ *is* compatible *with* μ, *if there exists* $\mathbf{M}_k \in \Lambda$ *such that*

1. $\mu \in \mathbf{M}_k$,
2. *The number of knots in* $\mathbf{M}_k$ *is equal to the number of knots in the minimal representation of* μ,
3. $p_{\mathbf{M}_k} < p_{\mathbf{M}_j}$ *for all* $j \neq k$ *such that* $\mu \in \mathbf{M}_j$.

In this case, we refer to $\mathbf{M}_k$ *as* compatible model.

Definition 4 *Define the following sequences*

$$(r(n))_{n\in\mathbf{N}} \text{ with } r^2(n) = \frac{n^2}{D^2(n)} \sim n^{\alpha}L(n),$$

$$(h(n))_{n\in\mathbf{N}} \subset \mathbf{R}_+^* \text{ such that } h(n) \to \infty.$$

If $(\xi_t)_{t\in\mathbf{Z}}$ *has either long or short memory, define*

$$(\lambda(n))_{n\in\mathbf{N}} \subset \mathbf{R}_+^* \text{ such that } \frac{\lambda(n)}{r^2(n)} \to 0 \text{ and } \frac{h(n)\log\log(n)}{\lambda(n)} \to 0.$$

If $(\xi_t)_{t\in\mathbf{Z}}$ *is antipersistent, define*

$$(\lambda(n))_{n\in\mathbf{N}} \subset \mathbf{R}_+^* \text{ such that } \frac{\lambda(n)}{n} \to 0 \text{ and } \frac{h(n)}{\lambda(n)} \to 0.$$

For example, we may choose $\lambda(n) = \log(n)$ and $h(n) = \sqrt{\frac{\lambda(n)}{\log\log(n)}}$.

Definition 5 *Let* $x(t) = \mu\left(\frac{t}{n}\right) + \xi_t$, $t = 1, \ldots, n$ *as above and let* $\Lambda = \{\mathbf{M}_1, \ldots, \mathbf{M}_q\}$ *be a compatible set of spline families. For* $i \in \{1, \ldots, q\}$, *define* $\tilde{\sigma}^2_{\mathbf{M}_{i,h}}$ *by*

$$\tilde{\sigma}^2_{\mathbf{M}_{i,h}} = \inf\left\{\frac{1}{n}\sum_{t=1}^{n}(x_t - f(t/n))^2 : f \in \mathbf{M}_i\left(h^{-1}(n)\right)\right\}.$$

The determination of $\tilde{\sigma}^2_{\mathbf{M}_{i,h}}$ can be regarded as a *restricted* least square estimation: given an observed series $x(1), \ldots, x(n)$, we obtain the optimal fit $\hat{\mu} \in \mathbf{M}_i\left(h^{-1}(n)\right)$. The space of admissible functions is restricted by a lower bound $h^{-1}(n)$ for the distance between knots. Note that $h^{-1}(n) \to 0$, since $h(n) \to \infty$.

Now, the information criterion can be defined as follows:

Definition 6 *If* $(\xi_t)_{t\in\mathbf{Z}}$ *has either short or long memory, define*

$$\mathbf{I}(\mathbf{M}_i) = \log\left(\tilde{\sigma}^2_{\mathbf{M}_{i,h}}\right) + p_{\mathbf{M}_i}\frac{\lambda(n)}{r^2(n)}.$$

If $(\xi_t)_{t\in\mathbf{Z}}$ *is antipersistent, define*

$$\mathbf{I}(\mathbf{M}_i) = \log\left(\tilde{\sigma}^2_{\mathbf{M}_{i,h}}\right) + p_{\mathbf{M}_i}\frac{\lambda(n)}{n}.$$

Given a set a models $\{\mathbf{M}_1, \ldots, \mathbf{M}_q\}$, we estimate the true model by minimising $\mathbf{I}(\mathbf{M}_i)$. This estimation procedure is consistent in the following sense:

Theorem 1 *Let* $\Lambda = \{\mathbf{M}_1, \ldots, \mathbf{M}_q\}$ *be a compatible set of spline families and* $\mathbf{M}_k$ *the compatible model. Then, for all* $i \neq k$,

$$\mathbf{P}\left[\mathbf{I}(\mathbf{M}_i) > \mathbf{I}(\mathbf{M}_k)\right] \to 1 \text{ as } n \to \infty.$$

The proof of this theorem relies on a law of the itertated logarithm by Lai and Wei (1982).

Note that the denominator n^{-1} in the case of antipersistence is of the same order as $r^{-2}(n)$ under short memory (up to a slowly varying function).

REFERENCES

Beran, J. and Weiershäuser, A. (2010). On spline regression under Gaussian subordination with long memory. *Journal of Multivariate Analysis*, 102(2):315-335.

Beran, J., Weiershäuser, A., Galizia, C.G., Rein, J., Smith, B., and Strauch, M. (2011). Inference for calcium-imaging data based on spline regression with fractional residuals and errors-in-variables corrections. Preprint.

Feder, P.I. (1975). On Asymptotic Distribution Theory in Segmented Regression Problems– Identified Case. *The Annals of Statistics*, 3(1):49-83.

Gallant, A.R. (1974). The theory of nonlinear regression as it relates to segmented polynomial regression with estimated join points. *Mimeograph Series No. 925*, Institute of Statistics, North Carolina State University, Raleigh.

Lai, T.L. and Wei, C.Z. (1982). A Law of the Iterated Logarithm for Double Arrays of Independent Random Variables with Applications to Regression and Time Series Models. *The Annals of Probability* 10(2): 320-335.

Liu, J., Wu, S., Zidek, J.V. (1997) On segmented multivariate regression, *Statistica Sinica*, 7(2):497-525.

Kim, J. and Kim, H.J. (2008). Asymptotic results in segmented multiple regression. *Journal of Multivariate Analysis*, 99(9):2016-2038.

Yao, Y.C. (1988) Estimating the number of change-points via Schwarz' criterion. *Statistics and Probability Letters* 6:181-189.