LTCC Advanced Course: Introduction to Semiparametric Modelling Lecture 4

Clifford Lam

Department of Statistics
London School of Economics and Political Science

Estimation of derivatives

- In local polynomial regression, the first derivative is the second entry of the estimated coefficients parameter $\hat{\beta}$. Bias and variance of the estimator have been presented in lecture 2. All advantages of local polynomial fitting retains for the derivatives estimations.
- For penalized polynomial splines, we use higher degree polynomial basis functions to ensure smoothness. For first derivative estimation, quadratic splines are the simplest basis leading to smooth fits.
- Let $\hat{f}$ be a quadratic penalized fit:

$$\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \sum_{k=1}^{K} \hat{u}_k (x - \kappa_k)_+^2.$$

The derivative estimate is then

$$\hat{f}'(x) = \hat{\beta}_1 + 2\hat{\beta}_2 x + \sum_{k=1}^{K} 2\hat{u}_k (x - \kappa_k)_+.$$

Estimation of derivatives

- If we let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)^T$ and $\hat{\mathbf{u}} = (\hat{u}_1, \ldots, \hat{u}_K)^T$, with

$$\mathbf{X}_x = (0, 1, 2x) \quad \text{and} \quad \mathbf{Z}_x = (2(x-\kappa_1)_+, \ldots, 2(x-\kappa_K)_+),$$

  then

$$\hat{f}'(x) = \mathbf{X}_x \hat{\boldsymbol{\beta}} + \mathbf{Z}_x \hat{\mathbf{u}}.$$

- The variance is

$$\begin{aligned} \text{var}\{\hat{f}'(x) - f'(x)\} &\approx \mathbf{C}_x \text{cov} \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\mathbf{u}} - \mathbf{u} \end{bmatrix} \mathbf{C}_x^T \\ &= \sigma^2 \mathbf{C}_x \left(\mathbf{C}^T \mathbf{C} + \frac{\sigma^2}{\sigma_U^2} \mathbf{D} \right)^{-1} \mathbf{C}_x^T, \end{aligned}$$

  where $\mathbf{C}_x = (\mathbf{X}_x, \mathbf{Z}_x)$ and $\mathbf{D} = \text{diag}(0, 0, 0, 1, \ldots, 1)$.

- Inference can be made using asymptotic normality of the estimated parameters,

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\mathbf{u}} - \mathbf{u} \end{pmatrix} \overset{\text{approx}}{\sim} N \left\{ \mathbf{0}, \hat{\sigma}^2 \left(\mathbf{C}^T \mathbf{C} + \frac{\sigma^2}{\sigma_U^2} \mathbf{D} \right)^{-1} \right\}.$$

Inference - example

- For the LIDAR data example, $-f'(\texttt{range})$ is proportional to the concentration of mercury at a given value of `range`. We estimate the derivative using a 15-knot penalized cubic spline.
- The tuning parameter is chosen by GCV, which minimizes the bias around the bump, where the estimate is significantly positive. This bump reveals a plume of mercury.
- Since GCV minimizes the bias around such bump, a global tuning parameter λ is chosen to be quite small. This small λ then produces unnecessary wiggles on those `range` values outside of the bump, which should remain zero.
- An **spatially adaptive** λ is preferable in this case.

Inference - example

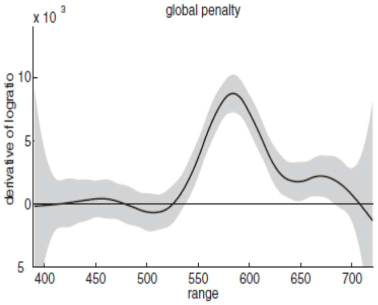

Figure: LIDAR data: estimate of first derivative with pointwise CIs, using global penalty.

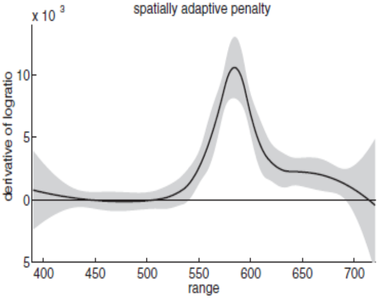

Figure: Same, but spatially adaptive penalty is used.

Semiparametric binary offset model

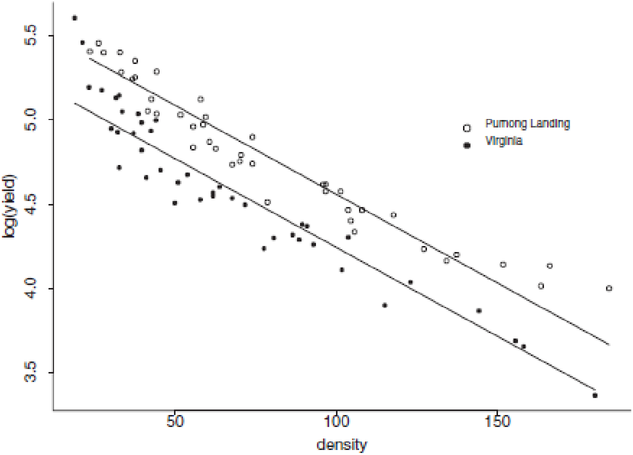

Figure: Scatterplot of the density and log.yield for the onions data. The plotting symbols indicate the two locations where the onions were cultivated. The lines correspond to the linear additive model fit to the data.

Semiparametric binary offset model

- The onions data is on yields (g/plant) of white Spanish onions in two locations: Purnong Landing and Virginia, South Australia. The horizontal axis corresponds to areal density of plants (plants/m^2).
- We model by
$$\log(\texttt{yield}_i) = \beta_0 + \beta_1 \texttt{PL}_i + \beta_2 \texttt{density}_i + \epsilon_i,$$
where
$$\texttt{PL}_i = \begin{cases} 0, & \text{if } i\text{th measurement is from Virginia;} \\ 1, & \text{if } i\text{th measurement is from Purnong Landing.} \end{cases}$$
- Since slight curvature is apparent in the scatterplots for each location, we can generalize the model to
$$\log(\texttt{yield}_i) = \beta_1 \texttt{PL}_i + f(\texttt{density}_i) + \epsilon_i.$$
- The model is called a semiparametric binary offset model. It has both parametric and nonparametric components. The variable `PL` vertically offsets the relationship between $E(\log(\texttt{yield}_i))$ and `density` according to location.

Semiparametric binary offset model

- The penalized linear spline formulation is

$$\log(\texttt{yield}_i) = \beta_0 + \beta_1 \texttt{PL}_i + \beta_2 \texttt{density}_i + \sum_{k=1}^{K} u_k (\texttt{density}_i - \kappa_k)_+ + \epsilon_i,$$

  where $u_k \sim i.i.d. N(0, \sigma_U^2)$, $\epsilon_i \sim i.i.d. N(0, \sigma^2)$.

- The estimated location effect is

$$\hat{\beta}_1 = 0.3331, \quad \widehat{\mathsf{SD}}(\hat{\beta}_1) = 0.0239.$$

  A 95% confidence interval for it is $(0.286, 0.380)$. It is 77% of the length of that obtained from a model with `density` entering linearly.

Semiparametric binary offset model

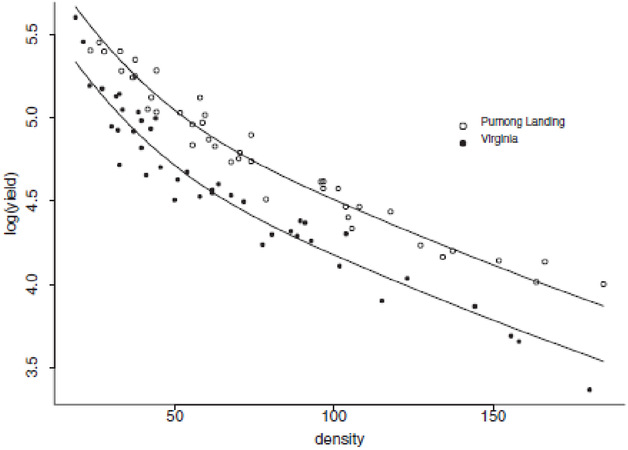

Figure: Onion data: fit to the additive model. The response is log.yield. The effect of density is fit by a penalized quadratic spline using REML to select the penalty parameter.

Interactions

- In reality, the effect of location may change according to the value of `density`. Likewise, the effect of `density` may change also depending on location.
- The additive model we considered has **fixed difference** between locations as `density` changes. Likewise, at different locations, changes in log(`yield`) is the same as `density` changes from a certain value to another. Hence it is not flexible enough.
- We consider **interactions** between the two factors. The most general interaction can be

$$E\{\log(\texttt{yield})_i\} = \begin{cases} f_{\mathsf{PL}}(\texttt{density}_i), & \text{if Purnong Landing;} \\ f_{\mathsf{VA}}(\texttt{density}_i), & \text{if Virginia.} \end{cases}$$

- To fit the model we treat the locations separately. But in the end the fits look similar to the additive model, suggesting additivity is a reasonable assumption.

Semiparametric binary offset model

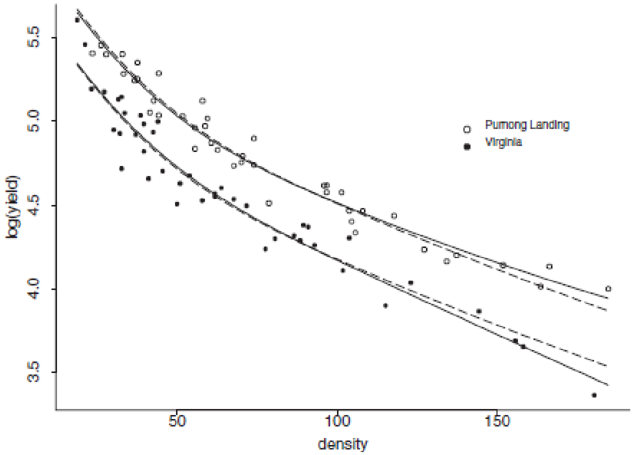

Figure: Additive and interaction fits to onion data. The solid curves are REML-based fits of the general interaction model. The dashed curves are REML-based fits of the additive model.

Testing for additivity

- Under general interactions between `location` and `density`, we are fitting

$$y_i = f_{z_i}(x_i) + \epsilon_i,$$

where $(x_i, y_i, z_i) = (\texttt{location}_i, \log(\texttt{yield}_i), \texttt{density}_i)$, and

$$z_i = \begin{cases} 1, & \text{if } (x_i, y_i) \text{ is from Purnong Landing;} \\ 2, & \text{if } (x_i, y_i) \text{ is from Virginia.} \end{cases}$$

- Using a K-knots p-th degree polynomial spline, the additive model can be written

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_i^j + \sum_{k=1}^{K} u_k (x_i - \kappa_k)_+^p + \gamma_0 z_i + \epsilon_i.$$

- The model with interactions is

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_i^j + \sum_{k=1}^{K} u_k (x_i - \kappa_k)_+^p$$

$$+ \sum_{\ell=1}^{2} z_{i\ell} \left(\gamma_{0\ell} + \sum_{j=1}^{p} \gamma_{j\ell} x_i^j + \sum_{k=1}^{K} v_{k\ell} (x_i - \kappa_k)_+^p \right) + \epsilon_i,$$

where $z_{i\ell} = 1$ if $z_i = \ell$ and 0 otherwise. We also constraint $v_{k\ell} = 0$ for $\ell = 1$.

- Assume $u_k \sim i.i.d. N(0, \sigma_u^2)$ and $v_{k\ell} \sim i.i.d. N(0, \sigma_v^2)$ (mixed model).

Testing for additivity

- With the two models in spline formulation, the simple additive model assumes that all $\gamma_{j\ell} = 0$ for $j = 1, \ldots, p$ and $\gamma_{01} = 0$, $\ell = 1, 2$, and the $v_{k\ell}$'s are all 0.
- In mixed model formulation, it means that

$$H_0 : \gamma_{01} = 0, \gamma_{i\ell} = 0, \;\; j = 1, \ldots, p, \; \ell = 1, 2, \;\;\; \sigma_v^2 = 0.$$

- Since we overparametrized the model with interaction, we need to count the number of restrictions needed. In our example, we can constraint for example $\gamma_{j\ell} = 0$ for $\ell = 2$. Hence there are p restrictions.
- The number of restrictions we placed on the fixed component is then $2p + 1 - (p + 1) = p$.
- We can use the likelihood ratio test to test this hypothesis. Null distribution of the statistic can be simulated (preferred), or we can use the asymptotic approximation which is an even mixture of chi-square distributions.

Additive model

- In almost all examples and theories developed so far, we have dealt with only one predictor variable. Yet most of the times there will be more than one predictor variables.
- It is easy to generalise what we have learnt to multiple smooth functions, one for each predictor variable. The effects are added together, and so it is called an additive model.
- Consider a data example where 56 US cities are observed with minimum temperature y and geographical locations (degree latitude s, degree longitude t). A nonparametric additive model to consider is then

$$y_i = \beta_0 + f(s_i) + g(t_i) + \epsilon_i,$$

  where f and g are smooth functions.
- We can easily extend our penalized linear spline formulation for additive model:

$$y_i = \beta_0 + \beta_s s_i + \sum_{k=1}^{K_s} u_k^s (s_i - \kappa_k^s)_+ + \beta_t t_i + \sum_{k=1}^{K_t} u_k^t (t_i - \kappa_k^t)_+ + \epsilon_i,$$

  where κ_k^s represents knots in the s direction, and similar definitions for κ_k^t.

Additive model

- The vector of fitted model is

$$\hat{\mathbf{y}} = \mathbf{C}(\mathbf{C}^T\mathbf{C} + \Lambda)^{-1}\mathbf{C}^T\mathbf{y},$$

where $\Lambda = \text{diag}(0,0,0,\lambda_s^2,\ldots,\lambda_s^2,\lambda_t^2,\ldots,\lambda_t^2)$, and

$$\mathbf{C} = \left[1, s_i, t_i, (s_i - \kappa_1^s)_+, \ldots, (s_i - \kappa_{K_s}^s)_+, (t_i - \kappa_1^t)_+, \ldots, (t_i - \kappa_{K_t}^t)_+\right]_{1 \le i \le n}.$$

- This model can also be fitted using the mixed model formulation. Let $\boldsymbol{\beta} = (\beta_0, \beta_s, \beta_t)^T$, $\mathbf{u} = (u_1^s, \ldots, u_{K_s}^s, u_1^t, \ldots, u_{K_t}^t)^T$, and

$$\mathbf{X} = [1, s_i, t_i]_{1 \le i \le n},$$
$$\mathbf{Z} = \left[(s_i - \kappa_1^s)_+, \ldots, (s_i - \kappa_{K_s}^s)_+, (t_i - \kappa_1^t)_+, \ldots, (t_i - \kappa_{K_t}^t)_+\right]_{1 \le i \le n},$$

then penalized least squares is equivalent to BLUP in the mixed model $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$, with

$$E\begin{pmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{pmatrix} = \mathbf{0}, \quad \text{cov}\begin{pmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{pmatrix} = \begin{pmatrix} \sigma_s^2\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_t^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma^2\mathbf{I} \end{pmatrix},$$

and $\lambda_s = \sigma/\sigma_s$ and $\lambda_t = \sigma/\sigma_t$.

Additive model

- Hence effectively, using REML to fit the mixed model returns the two estimated tuning parameters $\hat{\lambda}_s = \hat{\sigma}/\hat{\sigma}_s$ and $\hat{\lambda}_t = \hat{\sigma}/\hat{\sigma}_t$. With more additive components, it become very apparent the advantages of using mixed model for estimation of parameters.
- Instead of plotting the estimated functions $\hat{f}$ and $\hat{g}$ alone, we usually plot
$$\hat{\beta}_0 + \hat{f}(\bar{s}) + \hat{g}(t)$$
as a function of t, and
$$\hat{\beta}_0 + \hat{f}(s) + \hat{g}(\bar{t})$$
as a function of s. This increases the usefulness of the plot as it shows the absolute contribution to the response when s or t changes.
- Pointwise variability bands are added as usual using
$$\text{cov}\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} - \mathbf{u} \end{pmatrix} = \sigma^2(\mathbf{C}^T\mathbf{C} + \Lambda)^{-1}.$$
- **Partial residuals** can be plotted. For the plot against s, the partial residuals are
$$y_i - \hat{\beta}_0 - \hat{g}(t_i) + \{\hat{\beta}_0 + \hat{g}(\bar{t})\} = y_i - \{\hat{g}(t_i) - \hat{g}(\bar{t})\}.$$
For the plot against t, the partial residuals are
$$y_i - \{\hat{f}(s_i) - \hat{f}(\bar{s})\}.$$

Additive model - Temperature example

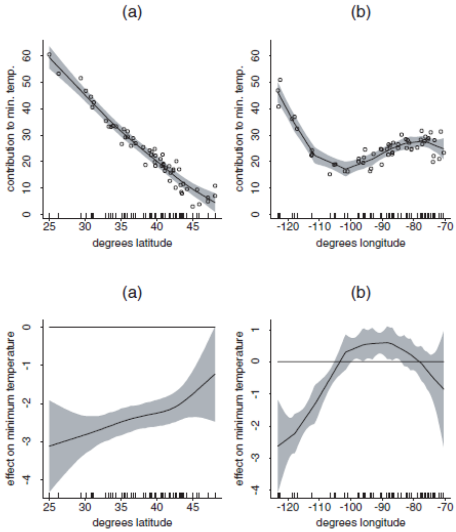

Figure: Upper row: Plot of component functions with the other variable fixed at mean value; partial residuals and variability bands added. Lower row: Derivative estimates with variability bands.