

1 chapter 2

1.1 time series

1.1.1 PLS

The basic model is

$$y_t = \gamma_t + \epsilon_t$$

where γ is the trend and ϵ are the zero-centered errors

We estimate γ by minimizing

$$\text{PLS}(\gamma) = \sum_{i=1}^T (y_t - \gamma_t)^2 + \lambda \sum_{i=3}^T (\Delta \gamma)^2$$

If we write

$$\mathbf{D} = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \dots & & \\ & & & -1 & 1 \end{pmatrix}$$

Then

$$\text{pen} = \gamma' \mathbf{D}' \mathbf{D} \gamma = \gamma' \mathbf{K} \gamma$$

The usual manipulations give

$$\hat{\gamma}_{\text{PLS}} = (\mathbf{I} + \lambda \mathbf{K})^{-1} \mathbf{y}$$

In turn, this gives

$$\begin{aligned} E[\hat{\gamma}_{\text{PLS}}] &= (\mathbf{I} + \lambda \mathbf{K})^{-1} \gamma \\ \text{Cov}[\hat{\gamma}_{\text{PLS}}] &= \sigma^2 (\mathbf{I} + \lambda \mathbf{K})^{-2} \end{aligned}$$

If the errors are assumed Gaussian, so is $\hat{\gamma}_{\text{PLS}}$

1.1.2 Bayesian

The sampling distribution is

$$\mathbf{y} | \gamma \sim N(\gamma, \sigma^2 \mathbf{I})$$

The prior is the solution to the equation

$$\gamma_t - 2\gamma_{t-1} + \gamma_{t-2} = u_t$$

with $u_t \sim N(0, \tau^2)$. But additionally we have to give priors to the initial conditions, for instance

$$(\gamma_1, \gamma_2)' = N(0, c \mathbf{I}_2)$$

for $c \rightarrow \infty$, the prior becomes

$$p(\gamma) \propto \exp \left(-\frac{1}{2\tau^2} \gamma' \mathbf{K} \gamma \right)$$

In words, the penalty term is the negative log-prior

It turns out we have conjugacy, so the posterior is also normal. It's easy to derive that $\hat{\gamma}_{\text{PLS}}$ is identical to the posterior mean:

$$E(\boldsymbol{\gamma}|\mathbf{y}) = (\mathbf{I} + \lambda\mathbf{K})^{-1}\mathbf{y}$$

we also can obtain

$$\text{Cov}(\boldsymbol{\gamma}|\mathbf{y}) = \sigma^2(\mathbf{I} + \lambda\mathbf{K})^{-1}$$

The results hold for any \mathbf{K} . That is, it doesn't matter the order at which you penalize.

1.1.3 Modifications and extensions

What if you don't know σ and τ ? Either you treat them as parameters and you estimate them as any other (FB) or you use some frequentist technique like MLE (EB).

The model can be extended easily in two ways:

$$y_t = \gamma_t + \xi_t + \epsilon_t$$

where η_t is a seasonal component. The prior for that is given by an autoregressive series with *per* many terms. This also turns out to be Gaussian, with an appropriate penalty matrix.

Additional covariates can also be included:

$$y_t = \gamma_t + \xi_t + \mathbf{x}'_t\boldsymbol{\beta} + \mathbf{z}'_t\boldsymbol{\alpha}_t + \epsilon_t$$

The assumption of independent errors is questionable, so we can always model something like

$$\epsilon_t = \phi\epsilon_{t-1} + v_t$$

where $v_t \sim N(0, \sigma^2)$ and $|\phi| < 1$.

If the trend is not smooth, we change the formulation of the prior for γ_t :

$$\gamma_t = 2\gamma_{t-1} - \gamma_{t-2} + u_t$$

We have two options here. Either we make $u_t|\omega_t \sim N(0, \tau^2/\omega_t)$ where $\omega_t \sim \chi^2_\nu$. Or we write $\tau = \tau_t$ and give it some stochastic process characteristics.

If the time intervals differ in length, we basically just reweight.

1.1.4 Non-Gaussian observation models

What if the observations are binary or binomial or counts?

If it is binary, we write

$$\gamma_t = \log\left(\frac{\pi_t}{1 - \pi_t}\right)$$

and then

$$\boldsymbol{\gamma}_{\text{pen}} = \arg \max_{\boldsymbol{\gamma}} l_{\text{pen}}(\boldsymbol{\gamma})$$

where

$$l_{\text{pen}}(\boldsymbol{\gamma}) = l(\boldsymbol{\gamma}) - \frac{\lambda}{2}\boldsymbol{\gamma}'\mathbf{K}\boldsymbol{\gamma}$$

$l(\boldsymbol{\gamma})$ is the log-likelihood of the binomial observations and \mathbf{K} is the corresponding penalty matrix.

We can proceed similarly for other distributions, obtained the so-called dynamic generalized linear models.

For a Bayesian formulation, we proceed in the same way as for the Gaussian models. We obtain the same correspondence between the penalized smoother and the posterior mode. But this cannot be computed exactly, we have to do numerical algorithms.

1.2 Semiparametric regression based on penalized splines

One problem with the models in the previous subsection is their high dimensionality, which makes them computationally expensive.

1.2.1 Gaussian observations

B-Splines Iterative definition of B-splines. Properties: local basis, unity decomposition and bounded range.

Regression The scatterplot smoothing problem becomes the linear problem

$$\mathbf{y} = \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$$

where

$$\mathbf{X} = \begin{pmatrix} B_1(x_1) & \cdots & B_K(x_1) \\ \vdots & & \vdots \\ B_1(x_n) & \cdots & B_K(x_n) \end{pmatrix}$$

The solution is just the OLS solution.

P-splines Just as we did before, we can penalize the coefficients to obtain the criterium

$$\text{PLS} = (\mathbf{y} - \mathbf{X}\boldsymbol{\gamma})'(\mathbf{y} - \mathbf{X}\boldsymbol{\gamma}) + \lambda\boldsymbol{\gamma}'\mathbf{K}\boldsymbol{\gamma}$$

For TPS, \mathbf{K} penalizes the coefficients associated with the local polynomials. For B-splines, it penalizes it's just as in the previous section.

An exact solution for $\hat{\boldsymbol{\gamma}}$ can be found by the usual methods.

If the knots are not evenly spaces, you have to weight just as in the previoius section.

For B-splines, when $\lambda \rightarrow \infty$, the function estimate approaches a $(d-1)$ -th order polynomial on x , provided that the degree of the spline is larger or equal to the order of the difference penalty.

Bayesian P-splines Similar results as before.

For TPS, the penalty described above is equivalent to non-informative priors on the coefficients associated to the global polynomials and Gaussian priors on the ones associated to the local ones. Once again, the PLS criterium and the posterior mode coincide. An exact expression can be found for the posterior of $\boldsymbol{\gamma}$.

For B-splines, the prior is formulated by an autoregressive process. We can choose either a first or second order random walk. The joint prior ends up being partially improper, with

$$p(\boldsymbol{\gamma}) = \left(\frac{1}{\tau^2}\right)^{\text{rank}(\mathbf{K})/2} \exp\left(-\frac{1}{2\tau^2}\boldsymbol{\gamma}'\mathbf{K}\boldsymbol{\gamma}\right)$$

where \mathbf{K} is the differences one.

A little more can be said about the first order one. The sequential conditional is

$$\gamma_k | \gamma_{k-1}, \dots, \gamma_1 \sim N(\gamma_{k-1}, \tau^2)$$

which allows one to make statements about the behavior of the estimate when τ^2 aquires extreme values. Also,

$$\gamma_k | \gamma_{-k} \sim N\left(\frac{1}{2}(\gamma_{k-1} + \gamma_{k+1}), \frac{\tau^2}{2}\right)$$

(excluding the coefficients at the boundary)

Bayesian inference We actually don't know σ^2 or τ^2 , so we have to include them among the parameters to estimate. In FB, we use MCMC. The priors are $\text{IG}(a_\sigma, b_\sigma)$ and $\text{IG}(a, b)$. We obtain conditional conjugacy and therefore a Gibbs sampler can be implemented.