GAUSSIAN PROCESS PRIOR EXERCISE

The exercise is due at the Thursday, 10/25 lecture. However, to be sure that the matrix algebra underlying the computations and plots for parts 3-6 of the exercise is understood by all, we will discuss it at the Tuesday 10/23 lecture. I'll ask one or more people in the class to explain the matrix algebra, or else where the algebra stumped them. In other words, by Tuesday you should have thought through how to do the exercise, though no code or work with data is due by then.

We will examine the univariate relationship between the (per capita) GDP growth variable in the FLS data set and the GDP level variable (GDPsh560), allowing for nonlinearity. Growth regressions tend to find a negative coefficient on the lagged level, which suggests a tendency of the GDP level to revert to a common mean. Of course for the most part this literature looks at the coefficient on lagged level GDP in the context of a regression with several other explanatory variables, but to make this doable as an exercise, we will look just at the bivariate relationship.

(1) Estimate a regression of the growth variable y on the level x and a constant, and display the regression line on a scatter plot with growth on the vertical axis and level on the horizontal.
(2) Estimate a regression of y on x, x^2 and a constant, and add the resulting fitted regression function to your plot.
(3) Estimate a non-parametric regression line using a Gaussian process prior. The model is

$$y_i = f(x_i) + \varepsilon_i\,, \tag{1}$$

with data i.i.d. across i, $\varepsilon_i \sim N(0, \sigma^2)$ and f a draw from a Wiener process with $\mathrm{Var}\big(f(x_1) - f(x_2)\big) = \nu^2\,|x_1 - x_2|$. This makes the covariance matrix of $\{f(x_1), \dots, f(x_n)\}$ a matrix with $\nu^2 \min(x_i, x_j) + K$ in the (i,j)th position. A Wiener process is non-stationary, so its unconditional variance is not defined. To get a covariance matrix for the $f(x_i)$ we have to specify an unconditional variance for some $f(x)$. K can be thought of as the variance of $f(0)$. By picking it very large, we are asserting that we know little about the level of f, so the data can determine it. Of course prior covariance matrix for $\{y_i\}$ is the sum of the covariance matrix of $\{f(x_i)\}$ and $\sigma^2 I$. Initially, choose $K = 1000$, $\sigma^2 = .02^2$ and $\nu^2 = .02^2$, Calculate the projection of $f(x)$ at a grid of 100 equally spaced x values on the observed y values and add the resulting line to your plot. (Use different colors, if possible.)
(4) Using a flat prior, (i.e. treating the likelihood as the posterior pdf) , find the maximum of the posterior density for σ^2 and ν^2, keeping K fixed at 1000. Calculate and plot the projection of f on the y observations using these optimized parameters.

(5) Calculate $\pm$ one standard deviation error bands around the estimated f and plot them. This can be a separate plot. Use the conditional distribution of f given the optimized σ^2 and ν^2 parameters.
(6) Calculate the weights on y values that the model uses to construct the estimated f values at 10 of the x values for which you have calculated an estimated f. These are implicit "kernels". Plot all 10 of these kernels on the same plot.

Note: The computations here are fairly straightforward once you have figured out what the appropriate matrix algebra is. Keep in mind that the error bands you are calculating reflect uncertainty about f given y — y is fixed, not random, in these calculations.