Exercise Series 4

1. The dataset `bmw` is a time series of log returns of the BMW stock (business-daily, between June 1986 and March 1990). The log return is defined as follows:

$$X_t = \log\left(\frac{P_t}{P_{t-1}}\right),$$

where P_t is the stock price at time t. Log returns can be modelled by

$$X_t = \sigma_t \epsilon_t, \text{ where } \mathbf{E}[\epsilon_t] = 0, \mathrm{Var}(\epsilon_t) = 1, \tag{1}$$

ϵ_t independent of $\{X_s;\ s < t\}$, $\sigma_t^2 = v(X_{t-1})$, where $v: I\!R \mapsto I\!R^+$ is the so-called "volatility function". Thus, X_t depends on $\{X_s;\ s < t\}$ only through X_{t-1} (Markov-property).

The model can be fitted by nonparametric regression of the function v in

$$Y_t = X_t^2 = v(X_{t-1}) + \eta_t, \text{ where } \eta_t = \sigma_t^2(\epsilon_t^2 - 1)$$

is treated as error term. The data can be downloaded at

`http://stat.ethz.ch/Teaching/Datasets/bmw.dat`

to a file with name `bmw.dat`, say, and read into **R** by

```
bmwlr <- scan("bmw.dat")
```

`bmwlr` should be a vector of 1000 observations.

a) Compute $\mathbf{E}[X_t|X_{t-1}, X_{t-2}, \ldots]$, $\mathrm{Var}(X_t|X_{t-1}, X_{t-2}, \ldots)$, $\mathrm{Cov}(X_t, X_{t-h})$, $h > 0$.
Background about conditional expectations:
For two (possibly multi-dimensional) random variables X and Y, the conditional distribution $P^{Y|X=x}$ can be uniquely defined for $P-$almost all values of X. Define $h(x) = \mathbf{E}[Y|X=x]$ as the expectation of Y under the conditional distribution $P^{Y|X=x}$. For the random variable X, $h(X) = \mathbf{E}[Y|X]$ is a random variable. Here is a very useful equation (the so-called tower property), which will be needed for parts a and b:

$$\mathbf{E}[Y] = \mathbf{E}[\mathbf{E}[Y|X]], \tag{2}$$

the outer expectation taken over the distribution of X. Conditional variances and covariances are defined analogously.

b) Show $\mathbf{E}[\eta_t] = 0$.
Note: Other usual model assumptions on errors, such as independence, are not fulfilled by η_t, but with some effort (don't try!) it can be shown that v can be optimally estimated by the same estimation methods as if the η_t would be independent errors.

c) Model (1) is often chosen for this kind of data because it leads to observations that are not autocorrelated[1] (as shown in part a), but dependent. Dependency can be verified by showing that under the model, $\mathrm{Cov}\left(X_t^2, X_{t-h}^2\right) \neq 0,\ h > 0$ (complicated). Plot and interpret the autocorrelation functions of X_t and X_t^2 for the BMW-dataset.

R-hint: Function `acf`. For example, "autocorrelation of lag 1" (in the plot, with 1000 observations, indicated as lag 1 out of 999) means correlation between X_t and X_{t-1}. The plot shows also an acceptance region (at 5%-significance level) for testing the null hypothesis of uncorrelated observations.

d) Fit the data using the nonparametric regression methods Nadaraya-Watson, local polynomial and smoothing splines for the regression function v. Check model assumptions. Comment on the results and compare the fits obtained using the previously mentioned nonparametric estimators.

R-hint: There are numerous functions to perform nonparametric regression and here are some of them. Consider the help-pages for details.

```
ksmooth       library(stats)       Nadaraya-Watson kernel regression
loess         library(stats)       local polynomial
smooth.spline library(stats)       smoothing splines
locpoly       library(KernSmooth)  local polynomial
dpill         library(KernSmooth)  optimal bandwith estimation
lokerns       library(lokern)      kernel regression with local
                                   optimal bandwith
```

Note that for computing residuals it is necessary to know the fitted values at the data points. For `ksmooth` they are provided via parameter `x.points` and for `loess` and `smooth.spline` via `fitted()`. For the other methods, a very dense grid has to be chosen and the nearest value in the grid could be used.

Warning: The structure in the data is pretty unclear. This is not a very didactic example.

Preliminary discussion: Friday, May 5, 2006.

Deadline: Friday, May 12, 2006, at the beginning of the seminar.

Advice: Every Thursday, 12.00-13.00, LEO C15, Leonhardstr. 27, or contact either Bernadetta Tarigan, `tarigan@stat.math.ethz.ch`, or Nicoleta Gosoniu, `gosoniu@ifspm.unizh.ch`.

[1] "Autocorrelated" refers to correlation over time, i.e., correlation between X_t and $X_{t-h},\ h > 0$.