Homework 2

Background:

- Make sure you have mastered homework 1. If necessary, go though the model answer posted on the website.
- Listen to Lecture on residuals, and review your notes from the lecture on temporal autocorrelation and physiological noise. Make sure you are familiar with Multivariate Gaussian Distribution and General Least Squares (see external lectures).

Data homework:

In this homework we will have a closer look at modelling of the low-frequency drifts in the time series. You will use the data set from the last homework. Additionally. we have now a T1-contrast image of the same slice, and the results of a probabilistic segmentation algorithm that tells us which voxels are likely gray matter and which voxels likely white matter. To find gray matter voxels, simply find voxels with GM>0.5. For white matter voxels, find voxels with WM>0.5.

T1	96x96	Axial slice through the anatomical image - T1 contrast.
GM	96x96	Probability of this voxel being gray matter.
WM	96x96	Probability of this voxel being white matter.

1. **Diagnosing the model**

   Start with your OLS solution from Homework 1. Make sure that you:
   a) ran the regression separately for each run
   b) you did not include the HPF regressors
   Use the residuals and beta weights from this estimation. Write a function that provides some diagnostics:

   a. Calculate the root mean square error for each all voxel (p) across all time points (t).

   $$RMS_p = \sqrt{\sum_{t=1}^{T}\left(y_{t,p} - \hat{y}_{t,p}\right)^2 / T}$$

   Calculate and report the mean RMS for all gray and white matter voxels. Which one is higher and why? (5pts)

   b. Plot the root-mean-square image. Where are there especially high residuals and why? (5pts)

   c. Compute the root-mean-square-timeseries across all gray matter voxels (GM>0.5). Before computing the root-mean-square time series, standardize the residual-time-series of each voxel by dividing it by the root-mean-square-error for that voxel, so each voxel has the same weight in this computation. (10pts)

   d. Using the residual time series, estimate and plot the auto-correlation function for temporal lags from t= 0 to 20, average for all gray-matter voxels and averaged for all white-matter voxels. What is the difference between gray and white matter? What could the reason for this difference be? (10pts)

   e. To estimate the variance of your estimators, calculate the consistency of the regression weights for the fingers k=1…10, across *gray matter voxels* p, across all runs r=1…8, where $\bar{b}_{pk}$ is the mean regression coefficient for voxel p, finger k (across runs):

$R^2 = 1 - RSS / TSS$

$TSS = \sum_p \sum_k \sum_r b_{pkr}^2$

$RSS = \sum_p \sum_k \sum_r \left(b_{pkr} - \overline{b}_{pk}\right)^2$

The coefficient can be interpreted as the amount of variance in the regression coefficients that is consistent across runs. (10pts)

2. **Improve the model by temporal filtering.**
   a. Remove the low frequency trends from the data by first regressing the data against the high-pass filter regressors (hpf). You can get the residuals by first calculating the residual-forming matrix for the high-pass filter regressors H = Xhpf:

   $$\mathbf{R} = \mathbf{I} - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T$$

   , where **I** an Identity matrix of size 984x984.
   The filtered data then is simply:
   $\tilde{\mathbf{Y}} = \mathbf{R}\mathbf{Y}$.
   Plot the same voxels as in homework 1. What do you notice? (10pts)

   b. Remove the low frequency trends from your task-related regressors (Xtask) and intercept by premultiplying R and plot the filtered and unfiltered first regressor from the first run. Plot 1 of the filtered task regressors as a line plot (for one run). How is the regressor changed through the filtering process? (10pts)

   c. Re-estimate the regression coefficients using filtered data and filtered design using OLS. Use your function from (1) to diagnose the residuals. What do you notice? Are the residuals and the consistency of the estimators improved? (10pts)

3. **General-least-squares estimation**
   Given the residuals, a good choice for the TxT variance-covariance matrix Φ for each run is a Toeplitz matrix, where the value of the nth diagonal is given by:
   $\phi_n = 0.7 * 0.98^n + 0.3$ and $\phi_0 = 1$.
   Generate and **plot t**his matrix. Then use it in a General-least-square (GLS) estimation.

   $$\hat{\mathbf{b}} = \left(\mathbf{X}^T\Phi^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\Phi^{-1}\mathbf{Y}$$

   Diagnose and compare the model quality using your diagnostic function from 1.
   Which one is better? What does this mean? (20pts)

4. Say we have a time-series model with

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \varepsilon$$

where y is a Tx1 vector and epsilon is i.i.d an (identical and independently distributed) vector with variance 1. Given a design matrix **X**, what is the expected variance-covariance matrix of your residuals from an OLS regression? (10pts)