Statistics 203: Introduction to Regression and Analysis of Variance

Functional Data

Jonathan Taylor

Today's class

- Functional data.
- A functional t-test.
- Mean, variance, covariance functions.
- Functional data I care about – fMRI.

Functional data

Functional data

- Last class, we talked about smoothing one function at a time.
  - B splines;
  - smoothing splines;
  - kernel smoothers.
- In a functional data setting we observe *many* functions at once, we might also observe many covariates for each curve.
- Example I will talk about later: functions are space-time images, covariates are simple: "motion correction" and "slow drift."

A two-sample functional t-test

- Suppose we observe a group of n paired curves: maybe the growth curves of twins separated at birth over the first 18 years of life but raised in different countries with a big difference in standard of living.
- Nutrition is known to have an effect on population height: twins' curves might be different.
- We can describe this data as $(Y_{i1,t}, Y_{i2,t}), 1 \leq i \leq n, 0 \leq t \leq 18$ where twins $Y_{\cdot 1}$ are in country # 1 and $Y_{\cdot 2}$ are in country # 2 with

$$Y_{ij,t} = \mu_{i,t} + \delta_t \cdot 1_{\{j=1\}} + \varepsilon_{ij,t}.$$

- The measurement noise ε can be assumed independent across subjects and set of twins put probably is dependent in time (μ_i would be a random effect curve for the i-th set of twins), and δ is the "nutrition effect"
- Although we never observe the *whole* curve, we should think of actually having an entire curve.

Functional t-test

- For each t, and each pair compute

$$\widehat{\delta}_t = \frac{1}{n}\sum_{i=1}^{n} Y_{i1,t} - Y_{i2,t},$$

- Also, compute

$$\widehat{\sigma}^2(\widehat{\delta}_t) = \frac{1}{n-1}\sum_{i=1}^{n}(Y_{i1,t} - Y_{i2,t} - \widehat{\delta}_t)^2$$

- Natural to use

$$T_t = \frac{\widehat{\delta}_t}{\widehat{\sigma}(\widehat{\delta}_t)}$$

  to test $H_{0,t} : \delta_t = 0$, i.e. at age t the "country effect" is 0.

- Test for no effect $H_0 = \cap_t H_{0,t}$: use

$$T_{\max} = \max_{t \in [0,18]} |T_t|.$$

Smooth noise

- Problem: what is the distribution of $T_{\max}$? It is not χ^2! Depends on properties of ε.
- Maybe we model each noise ε_{ij} as a smooth Gaussian process on $[0, 18]$.
- What is a Gaussian process? A random function such that for every collection $\{t_1, \dots, t_k\}$ the random vector

$$(\varepsilon_{t_1}, \dots, \varepsilon_{t_k})$$

  is multivariate normal.
- A Gaussian process ε_t is completely determined by

$$\mu_t = \mathbb{E}(\mu_t), \qquad R_{t,s} = \text{Cov}(\varepsilon_t, \varepsilon_s)$$

- If n is large (so T_t is almost Gaussian itself) then the distribution of

$$\varepsilon_{\max} = \max_{t \in [0,18]} \varepsilon_t$$

  depends on Var$(\dot{\varepsilon}_t)$ (known as Rice's formula).

Another approach

- Perhaps a more typical FDA approach would be "dimension" reduction. That is, take each curve $Y_{ij,t}$ and express it as a linear combination of basis functions $b_k(t)$ – the projection of $Y_{ij,t}$ onto the basis

$$Y_{ij,t} = \sum_k c_{ij,k} b_k(t) + r_{ij,t}$$

  Sometimes the remainder $r_{ij,t} = 0$ and no dimension reduction is used.
- In any case, once you have expressed each curve in a given basis, the two-sample t-test problem becomes a standard regression problem involving the coffi cients $c_{ij,k}$ which is the "new data."
- With this basis approach, it is possible to impose penalties (i.e. on the second derivatives, etc.) as long as you know how to compute the penalties in your specifi c basis.

Penalty example: second derivative

- Suppose we choose to express each curve $Y_{ij,t}$ as a linear combination of $\cos$ waves of the form

$$b_k(t) = \cos\left(2\pi k t/18\right), 1 \leq k \leq m.$$

- Then

$$b_k''(t) = c_k b_k(t), \qquad \int_0^{18} b_k(t) b_j(t) dt = \delta_{jk} c_k'.$$

- This means that

$$\int_0^T \left(\sum_k a_k b_k''(t) \right)^2 dt = \sum_k a_k^2 c_k^2 c_k'.$$

- Smoothing spline problem to estimate δ (penalty on integral of $\delta''(t)^2$) becomes a ridge problem.

Smoothing spline and ridge

- Suppose we look at the two-sample problem in the Fourier basis but we impose a second derivative penalty. Let

$$\widehat{Y}_{ij,t} = \sum_{k=1}^{m} c_{ij,k} b_k(t)$$

be the "dimension reduced" $Y_{ij,t}$'s and

$$\delta_t = \sum_k \delta_k b_k(t)$$

be a linear combination of sine waves.

- Problem find $\widehat{\delta}_t$ that minimizes

$$L_\lambda(\delta) = \sum_{i=1}^{n} \left(\widehat{Y}_{i1,t} - \widehat{Y}_{i2,t} - \delta_t \right)^2 \, dt + \lambda \int_0^{18} \delta''(t)^2 \, dt.$$

- Both integrals reduce to sums involving δ_k's and $c_{ij,k}$'s: equivalent to a ridge problem.

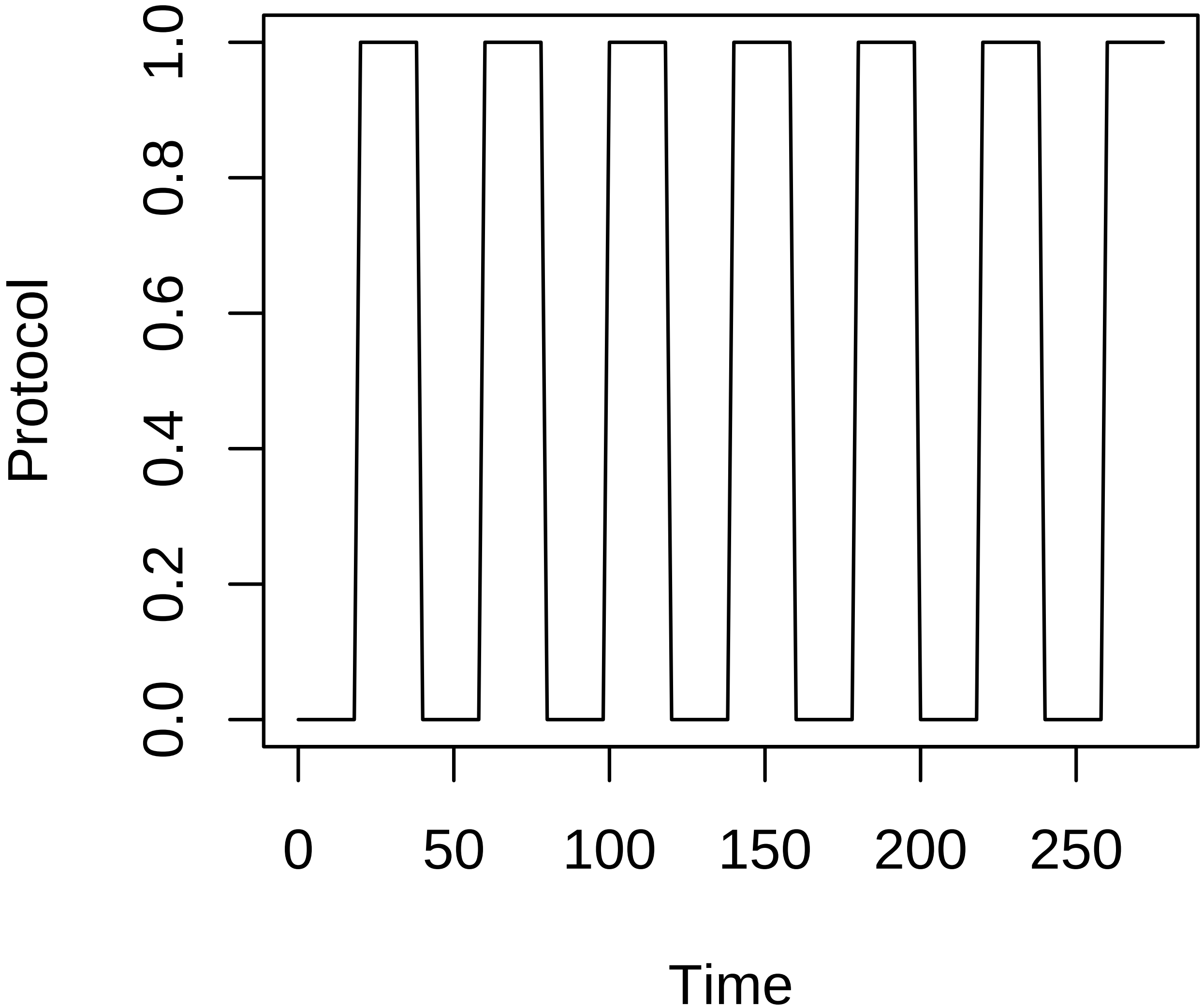
fMRI

What is fMRI?

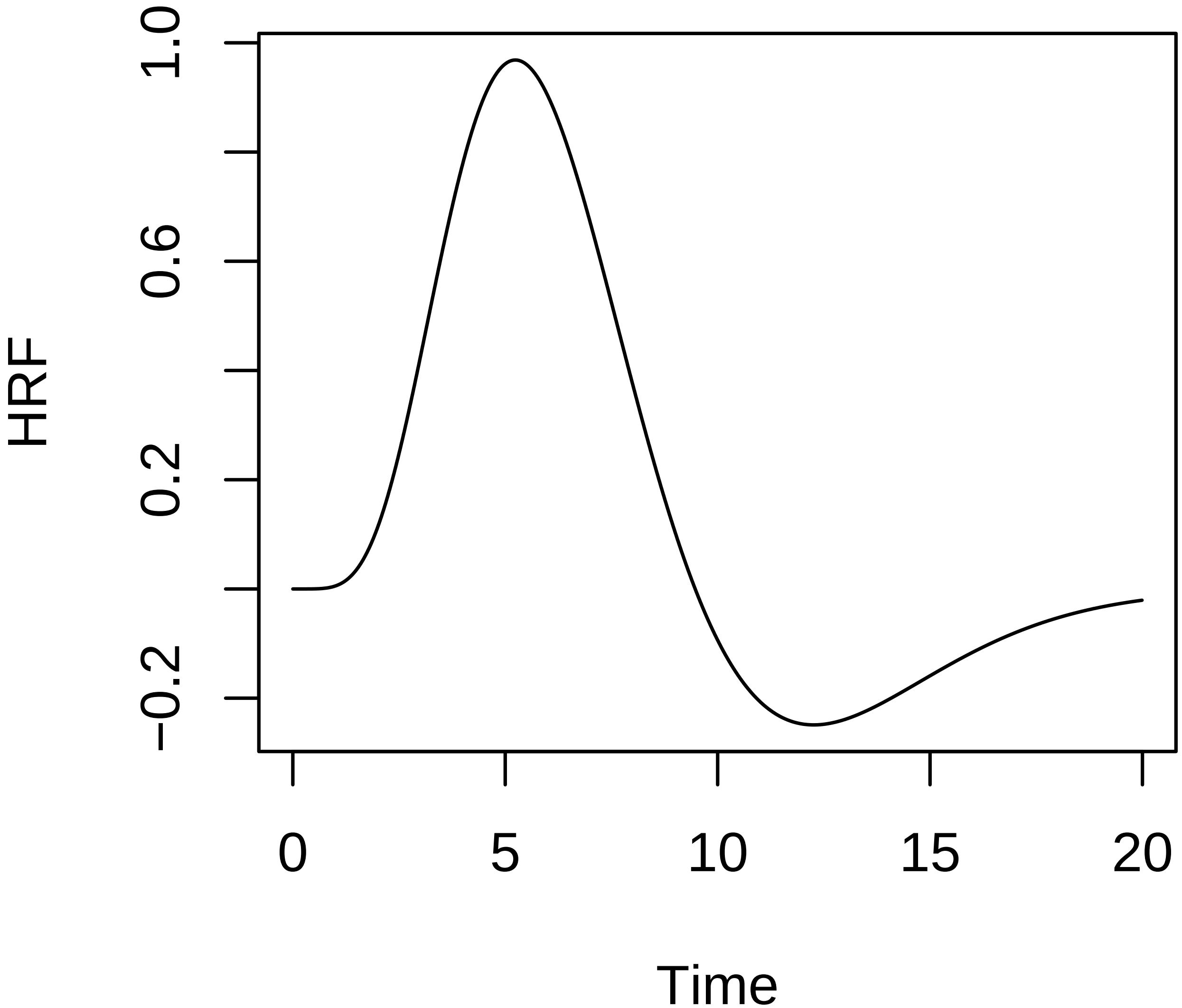
- fMRI (Functional Magnetic Resonance Imaging) is a space-time recording of "metabolic activity" in the human brain
- "paradigm" (motor task, i.e. fi nger tapping or cognitive task, i.e. face recognition) increases nerve cell activity in areas associated with the "paradigm"
- increased nerve cell activity increases metabolic demand for oxygen, increases metabolic activity results in a lagged increase in oxygenated Hg (hemoglobin)
- relationship between input "paradigm" and BOLD is modelled through a transfer function, the Hemodynamic Response Function (HRF)

$$\text{BOLD}_{x,t} = \beta_x^{\text{Input}} \cdot (\text{HRF} * \text{Input})_t + \sum_{i=1}^{k} \beta_{i,x} X_i + \text{Err}_{x,t}.$$

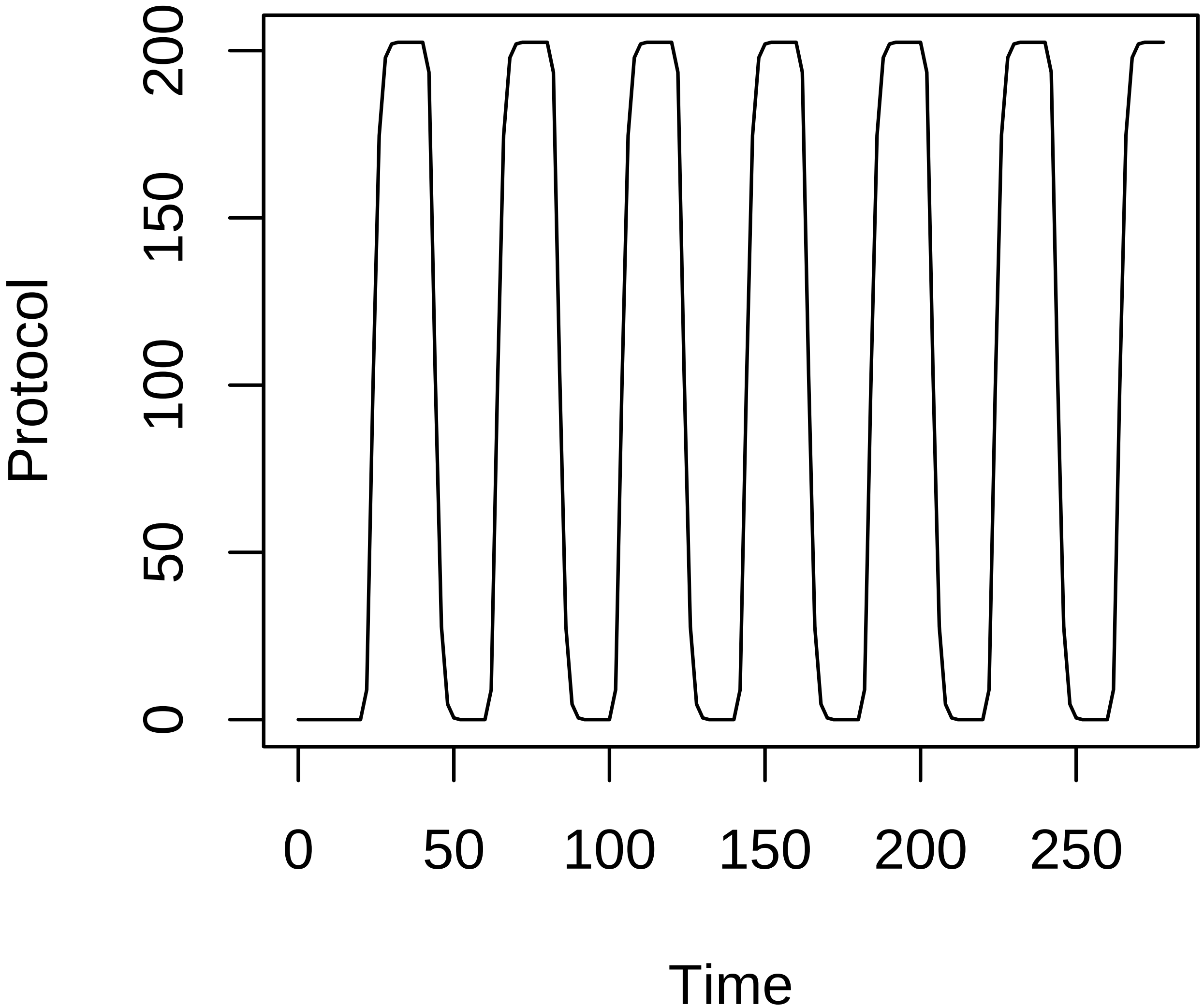
Block design – finger tapping



Hemodynamic response function



Convolved design – finger tapping



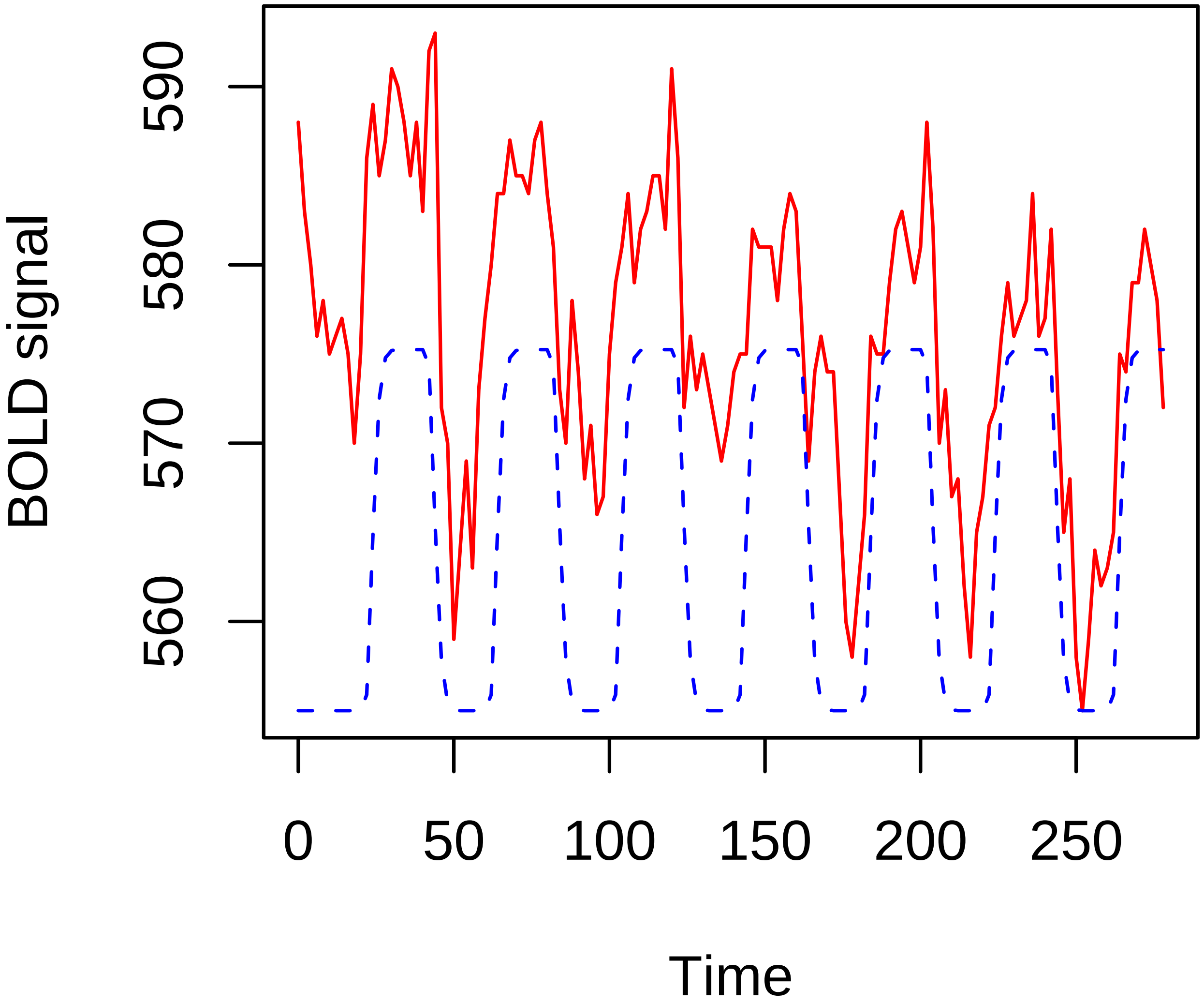
Components of Err$_{x,t}$

- physiological noise
  - cardiac noise
  - repiratory noise
  - basal metabolism
- motion artifacts
- saturation of signal
- other sources of error: lumped into "noise term" $\varepsilon_{x,t}$

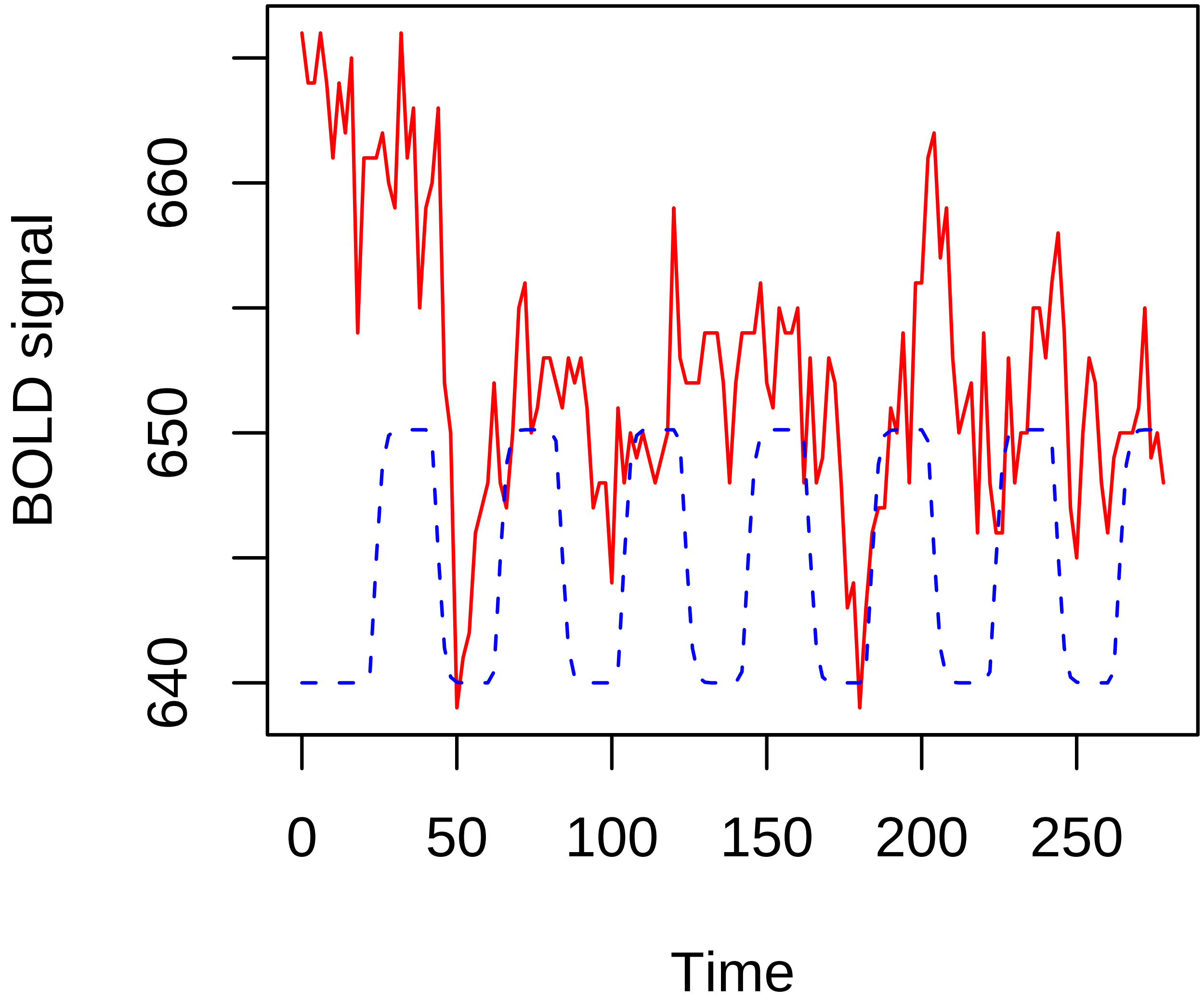
Full model for finger tapping data

- $$\text{BOLD}_{x,t} = \beta_x^{\text{Input}} \cdot (\text{HRF} * \text{Input})_t \; + \sum_{j=1}^{6} T_{j,t}\beta_{j,x}^{\text{Motion}} + \sum_{j=0}^{2} \beta_{j,x}^{\text{Time}} t^j + \varepsilon_{x,t}.$$
- Usually, model is fit voxel-by-voxel, usually spatial ignoring correlation between $\varepsilon_{x_1,\cdot}$ and $\varepsilon_{x_2,\cdot}$.
- Most common model is a two-stage procedure: first find a (smoothed) AR(1) coefficient image at each voxel in the brain: estimate standard errors and coefficients with this AR(1) value at each voxel.
  Sometimes, even correlation within time series at a single point is ignored and model is fit by OLS.
- Many people work on “improving” this basic model.

Voxel in the motor cortex



Marginally significant voxel



Real experiment: reward anticipation

- B. Knutson, my psychologist collaborator is interested in reward anticipation.
- Subjects play simple video game, and are told whether they can win, lose or draw on each round.
- If we contrast "win" (reward) vs. "draw" (neutral), we will fi nd areas that are sensitive to anticipating reward.

Combining subjects: fixed effect analysis

- For each subject i, compute

$$T_x^{RwAnt,i} = \frac{\widehat{\beta}_x^{RwAnt,i}}{\widehat{SE(\widehat{\beta}_x^{RwAnt,i})}}.$$

- Transform subject-specifi c comparisons to Tailarach space (common coordinates).
- Form group map

$$T_x = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} T_x^{RwAnt,i}.$$

- Analogous to the role of T_t in "growth curve" example.
- Where to set the threshold? The uncorrected 0.05 threshold is 1.96. The generalizations of Rice's formula set the threshold around 4.5 or so. Let's look at the difference.