Regression When the Predictors Are Images

Philip T. Reiss, New York University
Regression When the Predictors Are Images

Philip T. Reiss
New York University
phil.reiss@nyumc.org
http://works.bepress.com/phil_reiss

Wright State University
Dayton, Ohio
May 1, 2009

Joint work with Todd Ogden

Research supported in part by grant number 5F31MH073379-02, NIMH
Simple linear regression: fit least-squares line
Multiple linear regression:
fit least-squares (hyper)plane
Basic model equation
for multiple linear regression

\[ y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_{p-1} x_{i,p-1} + \varepsilon_i \]

for \( i = 1, \ldots, n \), where

- \( y_i \) is the \( i \)th subject’s outcome,
- \( \beta_0 \) is the intercept of the “true” line (or plane…),
- \( x_{i1}, \ldots, x_{i,p-1} \) are subject \( i \)’s values for predictor 1, \ldots, \( p - 1 \),
- \( \beta_1, \ldots, \beta_{p-1} \) are coefficients or effects of predictor 1, \ldots, \( p - 1 \),
- \( \varepsilon_i \) is an error term which in the “nicest” case is “iid normal”:
  1. independent across subjects,
  2. identically distributed for each subject,
  3. normally distributed.

Each of these 3 conditions can be relaxed.
• Above system of $n$ equations in $p$ unknowns can be written in inner product form as

$$y_i = x_i^T \beta + \varepsilon_i, \ i = 1, \ldots, n,$$

where $x = (1, x_{i1}, \ldots, x_{i,p-1})^T$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_{p-1})^T$, or in matrix form as

$$y = X \beta + \varepsilon$$

where $X$ is the $n \times p$ matrix with $i$th row $x_i$ (the “design matrix”).

• The least-squares estimate of $\beta$ is then

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \|y - X \beta\|^2.$$

• If rank$X = p$ then we have the unique minimizer

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

• But if $n < p$, there are ordinarily infinitely many $\beta \in \mathbb{R}^p$ such that $\|y - X \beta\|^2 = 0$. So we generally seek the optimal $\beta$ within some reasonable subset of $\mathbb{R}^p$. 
A motivating example: Serotonin receptors in depression

- Major depressive disorder affects $\approx 9.5\%$ of the U.S. population each year.
- Serotonin (5-HT), a neurotransmitter, is believed to play an important role in the disorder, mediated in part by distribution of 5-HT$_{1A}$ receptors in the brain.
- 5-HT$_{1A}$ receptor binding potential (BP), a measure of the receptors’ availability, can be mapped at each voxel (volume unit) of the brain via PET imaging and tracer kinetic modeling.
- Question: How can we relate these BP maps to a depression-related outcome such as Hamilton Depression Score (HAM-D)?
Idea: regress HAM-D (scalar) on BP (image)

\[ y_i = \alpha + s_i^T f + \varepsilon_i \]
• This can be thought of
  1. as super-high-dimensional regression,
  2. as regression with *functional data* ($L^2$ inner product)
• Many potential applications in psychiatry, including
  1. predicting treatment response in depression
  2. predicting progression from mild cognitive impairment to Alzheimer’s disease
**Road map**

<table>
<thead>
<tr>
<th></th>
<th>Linear regression</th>
<th>Generalized linear regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-D signal predictors</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2-D image predictors</td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>
An easier application: near-infrared spectroscopy

We model a vector $y$ of $n$ scalar responses as

$$y = X\alpha + Sf + \varepsilon$$

where

- $X$ is an $n \times p$ covariate matrix (may just be a column of 1s)
- $S$ is an $n \times N$ matrix:
  - $i$th row $s_i$ represents a signal predictor defined at points $v_1, \ldots, v_N$, and each column has mean zero
- $\varepsilon$ denotes iid errors

Since $n \ll N$, must reduce the dimension of $S$ to solve for coefficient function $f$. 
Previous approaches to solving for $f$:
1. Principal component regression (Massy, 1965)

Replace model $y = X\alpha + Sf + \varepsilon$ with

$$y = X\alpha + SV_q\beta + \varepsilon,$$

where $V_q$ comprises the $q$ leading columns of $V$, given the singular value decomposition $S = UDV^T$; take $\hat{f} = V_q\hat{\beta}$. 

2. Penalized $B$-spline expansion

Replace $y = X\alpha + Sf + \varepsilon$ with

$$y = X\alpha + SB\beta + \varepsilon$$

where the columns of $B$ form a $B$-spline basis; take $\hat{f} = B\hat{\beta}$, where $(\hat{\alpha}, \hat{\beta})$ minimize the penalized least squares criterion

$$\|y - X\alpha - SB\beta\|^2 + \lambda\beta^T P\beta.$$

- $P$ is chosen so that $\beta^T P\beta$ is a measure of the “roughness” of $f = B\beta$, e.g.,\[\int_{-\infty}^{\infty} f''(v)^2 dv\] (Cardot, Ferraty, and Sarda, 2003) or difference penalty (Marx and Eilers, 1999).

- The choice of $\lambda > 0$ governs the tradeoff between “fidelity to the data” and smoothness of the coefficient function (higher $\lambda \rightarrow$ smoother).
Reiss and Ogden (2007) propose to combine the above two approaches to obtain a coefficient function estimate

\[
\hat{f} = BV_q \hat{\zeta}
\]

where

- \(B\) is \(N \times K\) with columns forming a \(B\)-spline basis as before;
- \(V_q\) is a \(K \times q\) matrix whose columns are the leading columns of \(V\), given the singular value decomposition \(SB = UDV^T\).

This “functional principal component regression” (FPCR) modifies the above penalized least squares criterion

\[
\|y - X\alpha - SB\beta\|^2 + \lambda\beta^TP\beta
\]

by further restricting to the \(q\) leading principal components, to obtain

\[
\|y - X\alpha - SBV_q\zeta\|^2 + \lambda\zeta^TV_q^TPV_q\zeta.
\]
The second dimension reduction $SB \rightarrow SBV_q$ is optimal in the following minimax sense.

**Proposition 1.** Let $UDV^T$ be the singular value decomposition of the $n \times K$ matrix $Z$, let $U_q$ be the matrix consisting of the first $q < \min(n, K)$ columns of $U$, and let $D_q$ be the $q \times q$ upper left submatrix of $D$. Then $M_0 = U_q D_q$ minimizes

$$\max_{w \in \mathbb{R}^K, \|w\| = 1} \|Zw - \text{proj}_M Zw\|$$

over all $n \times q$ matrices $M$, where $\text{proj}_M$ denotes projection onto the column space of $M$.

In other words:
If we wish to replace an $n \times K$ design matrix $Z = SB$ with an $n \times q$ design matrix $M$ with $q < K$, the maximum perturbation in fitted values is minimized by taking $M$ to be $U_q D_q$, which equals $SBV_q$, the FPCR design matrix.
Tuning parameter 1: Number of principal components

- The $q$ in
  \[ \| \mathbf{y} - X\alpha - SBV_q\zeta \|_2^2 + \lambda \zeta^T V_q^T PV_q \zeta \]
  can be chosen by multifold cross-validation:
  1. Divide the data points $(y_i, x_i, s_i)$ into, say, 5 “validation sets.”
  2. For given $q$, and for $k = 1, \ldots, 5$,
     - remove the $k$th validation set and obtain estimates $\hat{\alpha}_{(-k)}(q)$, $\hat{\zeta}_{(-k)}(q)$ with the remaining data (the “training set”);
     - use those estimates to obtain predicted values $\hat{\mathbf{y}}_{k}^{(-k)}(q)$ for the left-out outcomes;
     - obtain prediction error $\| \mathbf{y}_k - \hat{\mathbf{y}}_{k}^{(-k)}(q) \|_2^2$ for the $k$th validation set.
  3. Choose $q$ such that $\sum_{k=1}^{5} \| \mathbf{y}_k - \hat{\mathbf{y}}_{k}^{(-k)}(q) \|_2^2$ is minimized.
- Alternatively, just choose $q$ that seems large enough to capture the necessary detail.
Tuning parameter 2: Roughness penalty parameter $\lambda$

Given $q$, the $\lambda$ in $\|\mathbf{y} - \mathbf{X}\alpha - SBV_q\zeta\|^2 + \lambda \zeta^T V_q^T PV_q \zeta$ can be chosen by optimizing over $\lambda$ a criterion function such as:

1. generalized cross-validation (GCV; Craven and Wahba, 1979)—related to cross-validation, but does not require fitting the model repeatedly.

2. restricted maximum likelihood (REML; Ruppert, Wand, and Carroll, 2003)—relies on connection with linear mixed models.
Reiss and Ogden (2009a) derived a 1-parameter family of functions $h_k (k \geq 1)$ s.t.

$$\text{sgn} \left[ \frac{d}{d\lambda} \text{REML}(\lambda) \right] = \text{sgn}[h_2(\lambda) - h_1(\lambda)] \text{ and}$$

$$\text{sgn} \left[ \frac{d}{d\lambda} \text{GCV}(\lambda) \right] = \text{sgn}[h_2(\lambda) - h_3(\lambda)].$$

In non-pathological cases, $h_1$ crosses $h_2$ (from smaller to larger) at a unique value $\hat{\lambda}_{REML}$, whereas $h_3$ crosses $h_2$ (from larger to smaller) at a unique point $\hat{\lambda}_{GCV}$. REML smooths more than GCV iff the latter crossing of $h_2$ precedes the former one.
Some asymptotic results

Assume the outcomes are generated by the model \( y = \alpha 1 + S^* B \beta + \varepsilon \) where

- \( S^* = (s_1^*, \ldots, s_n^*)^T \);
- \( s_i^* = (s_{i1}^*, \ldots, s_{iN}^*)^T, i = 1, 2, \ldots \), which are i.i.d. random vectors with \( E(s_{1j}^*) = 0 \) and \( E(s_{1j}^{*4}) < \infty \) for each \( j = 1, \ldots, N \);
- \( \varepsilon \) is a vector of i.i.d. errors with mean zero and finite variance, independent of \( S^* \); and
- \( B \) is a fixed \( N \times K \) B-spline basis matrix.

Let \( V_q^* \) be the \( K \times q \) population principal component matrix whose columns are the eigenvectors of \( E(B^T s_1^* s_1^{*T} B) \) corresponding to its leading eigenvalues \( \xi_1 > \ldots > \xi_q > 0 \), and let \( \Xi_q = \text{diag}(\xi_1, \ldots, \xi_q) \).

**Theorem 1.** Suppose

\[
\beta = V_q^* \zeta \text{ for some } \zeta = (\zeta_1, \ldots, \zeta_q)^T.
\]

If \( \hat{\beta}_n = V_q \hat{\zeta} \) denotes a \( q \)-component FPCR estimate for which \( \lambda_n \) is chosen to be \( o_p(n^{1/2}) \), then \( n^{1/2}(\hat{\beta}_n - \beta) \rightarrow_d Z_1 + Z_2 \), where \( Z_1 \sim N_K(0, \sigma^2 V_q \Xi_q^{-1} V_q^{*T}) \) and \( Z_2 \sim N_K(0, W) \) for a \( K \times K \) matrix \( W \) not depending on \( \sigma^2 \).
Under mild assumptions, the $\lambda_n = o_P(n^{1/2})$ condition imposed in Theorem 1 is met if $\lambda_n$ is chosen by GCV or REML. Indeed, a stronger condition then holds:

**Theorem 2.** Let $\lambda_n$ be the GCV or REML value associated with the $q$-component FPCR estimate. Assume that $V_q^* P V_q^*$ is nonsingular, and that $f = B \beta$ with $V_q^* B \beta \neq 0$. Then there exists $M > 0$ such that $P(\lambda_n > M) \to 0$ as $n \to \infty$. 
Consistency of choosing number of components by multifold CV:

**Theorem 3.** Assume that $f = BV^*_q \zeta$ with $\zeta_q \neq 0$. Suppose the number of FPCR components is chosen by multifold CV using $D_n$ divisions of the $n$ observations into training and validation sets of size $n_t$ and $n_v = n - n_t$, respectively, with $\lambda_n = o_P(n^{1/2})$. If $n_t, n_v \to \infty$ and $D_n = o_P[(\min\{n_t, n_v\})^{1/2}]$, then for any positive integer $q_1 < q$, the $q$-component model will be chosen over the $q_1$-component model with probability tending to 1 as $n \to \infty$. 
Simulation study

Given an \( n \times N \) signal matrix \( S \) (\( N \)-dimensional signal for each of \( n \) subjects), create a true coefficient function \( f \in \mathcal{R}^N \) and simulate outcomes \( y = Sf + \varepsilon \), where \( \varepsilon \) consists of \( n \) iid normal errors. Then use FPCR to derive estimate \( \hat{f} \). Two key performance metrics are

- estimation error \( \| \hat{f} - f \|^2 \),
- prediction error \( \| \hat{y} - E(y|S) \|^2 = \| S\hat{f} - Sf \|^2 \).
Table 1: Average Scaled Mean Squared Error of Prediction.

<table>
<thead>
<tr>
<th></th>
<th>Smooth function</th>
<th>Bumpy function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wheat</td>
<td>Gasoline</td>
</tr>
<tr>
<td></td>
<td>0.9 0.6</td>
<td>0.9 0.6</td>
</tr>
<tr>
<td>PBSE-GCV</td>
<td>.0065 .0327</td>
<td>.0127 .0703</td>
</tr>
<tr>
<td>PBSE-REML</td>
<td>.0050 .0178</td>
<td>.0095 .0528</td>
</tr>
<tr>
<td>PBSE-oracle</td>
<td>.0049 .0152</td>
<td>.0085 .0426</td>
</tr>
<tr>
<td>PCR</td>
<td>.0124 .0492</td>
<td>.0279 .1176</td>
</tr>
<tr>
<td>B-spline PCR</td>
<td>.0089 .0354</td>
<td>.0148 .0958</td>
</tr>
<tr>
<td>FPCR_C</td>
<td>.0094 .0334</td>
<td>.0200 .1029</td>
</tr>
<tr>
<td>FPCR_R-GCV</td>
<td>.0065 .0271</td>
<td>.0111 .0651</td>
</tr>
<tr>
<td>FPCR_R-REML</td>
<td>.0055 .0214</td>
<td>.0088 .0515</td>
</tr>
<tr>
<td>FPCR_R-oracle</td>
<td>.0063 .0211</td>
<td>.0097 .0608</td>
</tr>
<tr>
<td>FPCR_R-plug-in</td>
<td>.0088 .0338</td>
<td>.0129 .0824</td>
</tr>
<tr>
<td>PLS</td>
<td>.0135 .0497</td>
<td>.0278 .1070</td>
</tr>
<tr>
<td>B-spline PLS</td>
<td>.0081 .0308</td>
<td>.0149 .0945</td>
</tr>
<tr>
<td>FPLS_C</td>
<td>.0112 .0526</td>
<td>.0190 .1746</td>
</tr>
<tr>
<td>FPLS_R-GCV</td>
<td>.0072 .0281</td>
<td>.0128 .0750</td>
</tr>
<tr>
<td>FPLS_R-REML</td>
<td>.0057 .0203</td>
<td>.0091 .0511</td>
</tr>
</tbody>
</table>
Table 2: Mean of $L^2$ Norm of Error (Root Integrated Squared Error) in Estimating $f$.

<table>
<thead>
<tr>
<th></th>
<th>Smooth function</th>
<th>Bumpy function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wheat</td>
<td>Gasoline</td>
</tr>
<tr>
<td></td>
<td>0.9  0.6</td>
<td>0.9  0.6</td>
</tr>
<tr>
<td>PBSE-GCV</td>
<td>0.97  2.04</td>
<td>1.07  2.17</td>
</tr>
<tr>
<td>PBSE-REML</td>
<td>0.65  0.72</td>
<td>0.36  0.59</td>
</tr>
<tr>
<td>PBSE-oracle</td>
<td>0.42  0.67</td>
<td>0.33  0.48</td>
</tr>
<tr>
<td>PCR</td>
<td>1.01  1.16</td>
<td>0.81  1.01</td>
</tr>
<tr>
<td>$B$-spline PCR</td>
<td>0.98  1.14</td>
<td>0.77  1.27</td>
</tr>
<tr>
<td>FPCR$_C$</td>
<td>1.08  1.43</td>
<td>1.57  2.76</td>
</tr>
<tr>
<td>FPCR$_R$-GCV</td>
<td>0.80  1.01</td>
<td>0.53  0.73</td>
</tr>
<tr>
<td>FPCR$_R$-REML</td>
<td>0.66  0.75</td>
<td>0.29  0.42</td>
</tr>
<tr>
<td>FPCR$_R$-oracle</td>
<td>0.85  0.86</td>
<td>0.54  0.65</td>
</tr>
<tr>
<td>FPCR$_R$-plug-in</td>
<td>0.94  1.07</td>
<td>0.69  1.06</td>
</tr>
<tr>
<td>PLS</td>
<td>1.01  1.18</td>
<td>0.78  0.94</td>
</tr>
<tr>
<td>$B$-spline PLS</td>
<td>0.94  1.04</td>
<td>0.74  1.15</td>
</tr>
<tr>
<td>FPLS$_C$</td>
<td>1.11  1.51</td>
<td>0.82  1.46</td>
</tr>
<tr>
<td>FPLS$_R$-GCV</td>
<td>0.87  1.04</td>
<td>0.59  1.05</td>
</tr>
<tr>
<td>FPLS$_R$-REML</td>
<td>0.68  0.74</td>
<td>0.31  0.44</td>
</tr>
<tr>
<td></td>
<td>Linear regression</td>
<td>Generalized linear regression</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>1-D signal predictors</td>
<td><strong>1</strong></td>
<td><strong>2</strong></td>
</tr>
<tr>
<td>2-D image predictors</td>
<td><strong>3</strong></td>
<td></td>
</tr>
</tbody>
</table>
Generalized linear models

- Problem: The regression model \( y_i = x_i^T \beta + \varepsilon_i \) is not appropriate for certain types of outcome \( y_i \)—e.g., if \( y_i \) is binary (0/1).

- Under the standard assumptions that \( \varepsilon_i \) is independent of \( x_i \) and has expectation 0,

\[
\mu_i \equiv E(y_i|x_i) = x_i^T \beta.
\]

This motivates the generalized linear model

\[
\mu_i = g^{-1}(x_i^T \beta)
\]

where \( g \) is an invertible “link function.”

- Key example: **logistic regression.** If \( y_i \) is binary, \( \mu_i \) is simply \( p_i \equiv P(y_i = 1) \). With inverse link \( g^{-1}(t) = \frac{e^t}{1+e^t} \), the above becomes

\[
p_i = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}, \text{ i.e., } y_i \sim \text{Bernoulli}\left[\frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}\right].
\]
FPCR for generalized linear models

Reiss and Ogden (2009b) extended FPCR from the linear model
\[ \mu = X\alpha + S f \]

To the generalized linear model
\[ g(\mu) \equiv [g(\mu_1), \ldots, g(\mu_n)]^T = X\alpha + S f \]

Where \( g \) is an appropriate link function. Again we apply the restriction \( f = BV_q\zeta \) to obtain, e.g., the logistic model

\[ y_i \sim \text{Bernoulli} \left( \frac{\exp[(X\alpha + S BV_q\zeta)_i]}{1 + \exp[(X\alpha + S BV_q\zeta)_i]} \right). \]
<table>
<thead>
<tr>
<th></th>
<th>Linear regression</th>
<th>Generalized linear regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-D signal predictors</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2-D image predictors</td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>
Extension to image predictors

As for linear model with 1-D predictors, we seek to minimize

\[ \|y - X\alpha - SBV_q\zeta\|^2 + \lambda \zeta^T V_q^T P V_q \zeta, \]

but must modify \( B \) and \( P \).

- For the columns of \( B \), we chose radial cubic \( B \)-splines (Saranli and Baykal, 1998) centered at each of an equally spaced grid of knots \( \kappa_1, \ldots, \kappa_K \in \mathcal{R}^2 \).

- We chose \( P \) to yield the thin plate penalty given in two dimensions by

\[ \zeta^T V_q^T P V_q \zeta \approx \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ \left( \frac{\partial^2 f}{\partial v_1^2} \right)^2 + 2 \left( \frac{\partial^2 f}{\partial v_1 \partial v_2} \right)^2 + \left( \frac{\partial^2 f}{\partial v_2^2} \right)^2 \right] dv_1 dv_2. \]
Key scientific question re PET images:
Where in the brain (at which voxels $v$) is $f(v)$ significantly positive/negative?
• This question can be approached through simultaneous confidence bands for the coefficient image.

• A 95% (pointwise) confidence interval for \( f(v) \) is an interval \([L, U]\) (determined by the data) such that \( P(L \leq f(v) \leq U) = .95 \).

• A 95% simultaneous confidence band for \( f \) is given by functions \( \hat{f}_L(\cdot) \), \( \hat{f}_U(\cdot) \) such that \( P[\hat{f}_L(v) \leq f(v) \leq \hat{f}_U(v) \ \forall v] = .95 \).

• For a 1-D coefficient function this might look something like:
No analytic expression is available for $\hat{f}_L, \hat{f}_U$, so we pursue a bootstrapping approach (Mandel and Betensky, 2008):

1. Using, say, 999 random samples with replacement of $n$ data points $(y^*, x^*, s^*)$ from the $n$ observations, obtain bootstrap estimates $\hat{f}^*_1, \ldots, \hat{f}^*_{999}$ of the coefficient function.

2. At each $v$, order the function estimates: $\hat{f}^*_1(v) \leq \ldots \leq \hat{f}^*_{999}(v)$.

3. $[\hat{f}_{(25)}^*(v), \hat{f}_{(975)}^*(v)]$ is a pointwise 95% confidence interval for $f(v)$.

4. $\prod_v [\hat{f}_{(25)}^*(v), \hat{f}_{(975)}^*(v)]$ is in general not a simultaneous 95% confidence band for $f$; but the Cartesian product of, say, 99% pointwise intervals will be.

5. Given 95% simultaneous confidence bands $[\hat{f}_L, \hat{f}_U]$ for $f$, the coefficient function is declared significantly positive at all $v$ such that $\hat{f}_L(v) > 0$. 

32
Oversimplified example
More realistic example

- Function estimates cross over each other.
- Simultaneous band formed by Cartesian product of pointwise intervals from 2nd-smallest to 2nd-largest function:
Simulation studies

We used 68 maps of binding potential of 5-HT$_{1A}$ receptors obtained by Parsey et al. (2006) to perform two simulation studies:

**Study 1** Compare performance of FPCR with two other methods.

**Study 2** Assess how well the simultaneous inference procedure detects nonzero regions of $f$. 
True coefficient image $f$ for simulation studies
Simulation study 1: Performance comparisons

We compared three methods for both linear and logistic regression. The methods all begin with restricting the coefficient function to a spline basis, but differ in what happens next.

(a) **Roughness penalty only**—referred to above as the penalized $B$-spline expansion.

(b) **Reduction to leading PCs only**—similar to FPCR, but without a roughness penalty (i.e., $\lambda = 0$). Thus the only smoothing parameter is the number of components, which we chose by 5-fold cross-validation from among the values 1–10.

(c) **Leading PCs plus roughness penalty**—i.e., FPCR, with 35 PCs (accounting for 96% of the variation in $SB$).

For both linear and logistic regression, we generated 100 outcome vectors with equally spaced $R^2$ ("proportion of explained variation") values from .2 to .95, and computed each method’s prediction error.
Figure 1: (a) Linear regression results. (b) Logistic regression results.
Simulation study 2: Detecting significance

• For $R^2_L = .5, .6, .7, .8, .9$ (Menard, 2000), we simulated 20 sets of simulated binary outcomes

$$y_i \sim \text{Bernoulli}\left[\frac{\exp(s_i^T f)}{1 + \exp(s_i^T f)}\right]$$

$(i = 1, \ldots, 68)$, where

- $s_i$ denotes subject $i$’s binding potential map (1 slice, 5778 voxels), and
- $f = kf_0$, where $f_0$ is the artificial coefficient image shown previously and $k$ is chosen to attain the specified $R^2_L$.

• For each set of outcomes, we estimated $f$ by logistic FPCR with 35 PCs, and formed simultaneous confidence bands from 1999 bootstrap samples.

• Using these bands, we determined how many times (out of 20) each voxel was deemed significantly positive (or negative).
$f, \hat{f}$, and simultaneous bands: 10 examples with $R^2_L = .7$
$f$, $\hat{f}$, and simultaneous bands: zooming in
Discussion

• Regressing scalar outcomes on entire images is a very challenging problem, but our method seems to do quite well at detecting salient \((f \neq 0)\) regions.

• Extension to 3D images will require tackling several practical issues.

• Another extension: locally adaptive smoothing parameter.

• Ongoing work (Todd Ogden and Yihong Zhao) uses wavelets rather than splines, and seeks a sparse representation of the coefficient function.
Thank you!
References


