Web-Based Supplementary Materials for “Functional Generalized Linear Models with Images as Predictors”

Philip T. Reiss and R. Todd Ogden

August 28, 2008

A Proof of Proposition 1

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 $M_0 = U_qD_q$ minimizes

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

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

Proof. Since $\text{proj}_MZW = M(M^TM)^{-1}M^TZW$, (1) equals the spectral norm $\|Z - M(M^TM)^{-1}M^TZW\|_2$.

Extending the above notation, let $V_q$ consist of the first $q$ columns of $V$, let $U_{-q}$ consist of the last $K-q$ columns of $U$, and similarly define $D_{-q}$ and $V_{-q}$. We then have $Z = U_qD_qV_q^T + U_{-q}D_{-q}V_{-q}^T$, and thus

$$\text{proj}_{M_0}ZW = M_0(M_0^TM_0)^{-1}M_0^TZW$$

$$= U_qD_q(D_qU_q^TU_qD_q)^{-1}D_qU_q^T$$

$$\times (U_qD_qV_q^T + U_{-q}D_{-q}V_{-q}^T)w,$$

which simplifies, using the equalities $U_q^TU_q = I_q$ and $U_q^TU_{-q} = 0_{q \times (K-q)}$, to

$$\text{proj}_{M_0}ZW = U_qD_qV_q^T w.$$
But by a standard result (e.g., Watkins, 2002), \( \|Z - U_q D_q V_q^T\|_2 \leq \|Z - A\|_2 \) for any full-rank \( n \times q \) matrix \( A \), and in particular for any such \( A \) of the form \( M (M^T M)^{-1} M^T Z \) with \( M \) an \( n \times q \) matrix. This establishes the claim of minimality.

\[ \Box \]

B  Stretched Confidence Bands

This supplementary appendix concerns simultaneous confidence bands formed by bootstrap estimates \( \hat{f}^*_1, \ldots, \hat{f}^*_B \) of the coefficient function. In the paper we note that \( E(1) \), the band described by all \( B \) function estimates, may sometimes have estimated simultaneous coverage (in the cross-validatory sense described there) lower than \( 100(1 - \alpha)\% \). We describe here a method which can be used in such cases to “stretch” the envelope and thereby attain simultaneous coverage of \( 100(1 - \alpha)\% \).

B.1 Basic Formulation

A fairly straightforward stretching method is available under the condition that

\[ \text{for all } v, \quad \hat{f}^*_{(2)}(v) < \hat{f}(v) < \hat{f}^*_{(B-1)}(v). \tag{2} \]

For \( b = 1, \ldots, B \), let

\[ M_b = \max_v \max \left\{ \frac{\hat{f}^*_b(v) - \hat{f}(v)}{\hat{f}^*_{(2)}(v) - \hat{f}(v)}, \frac{\hat{f}^*_b(v) - \hat{f}(v)}{\hat{f}^*_{(B-1)}(v) - \hat{f}(v)} \right\}. \]

As argued in the paper, \( E(1) \) has estimated simultaneous coverage \( < 100(1 - \alpha)\% \) if and only if there are more than \( B \alpha \) values of \( b \in \{1, \ldots, B\} \) such that for some \( v = v_b \)

either \( \hat{f}^*_b(v_b) = \hat{f}^*_{(1)}(v_b) < \hat{f}^*_{(2)}(v_b) \) or \( \hat{f}^*_b(v_b) = \hat{f}^*_{(B)}(v_b) > \hat{f}^*_{(B-1)}(v_b). \)

Under (2), the above condition implies

\[ M_{\lceil B(1 - \alpha) \rceil} > 1. \tag{3} \]

(The left side of the above denotes the \( \lceil B(1 - \alpha) \rceil \)-th-smallest of the \( M_b \)'s, where \( \lceil B(1 - \alpha) \rceil \) is the smallest integer \( \geq B(1 - \alpha) \).) Thus \( E(1) \) will have at least \( 100(1 - \alpha)\% \) simultaneous coverage if (3) does not hold; the stretching procedure developed here is relevant only when (3) is observed to hold.

To define the stretched envelope, for each \( v \) let

\[ r(v) = \max \left\{ \frac{\hat{f}^*_1(v) - \hat{f}(v)}{\hat{f}^*_{(2)}(v) - \hat{f}(v)}, \frac{\hat{f}^*_B(v) - \hat{f}(v)}{\hat{f}^*_{(B-1)}(v) - \hat{f}(v)} \right\}, \tag{4} \]
and let
\[ c_\alpha(v) = \min \{ r(v), M_{(B(1-\alpha))} \}. \tag{5} \]

Note that, since \( r(v) \geq 1 \) for all \( v \), condition (3) implies
\[ c_\alpha(v) \geq 1 \text{ for all } v. \tag{6} \]

The stretched envelope is then given by
\[ E^{c_\alpha} = \prod_v \left[ \hat{f}(v) + c_\alpha(v)\{\hat{f}^*_v(v) - \hat{f}(v)\}, \hat{f}(v) + c_\alpha(v)\{\hat{f}^*_B(v) - \hat{f}(v)\} \right]. \tag{7} \]

The intuition behind this procedure is as follows. Equation (4) gives the smallest value \( r(v) > 0 \) such that if we instead defined the stretched envelope as
\[ \prod_v \left[ \hat{f}(v) + r(v)\{\hat{f}^*_v(v) - \hat{f}(v)\}, \hat{f}(v) + r(v)\{\hat{f}^*_B(v) - \hat{f}(v)\} \right], \]
then \( \hat{f}^*_b(v) \) would lie within the “delete-bth-function” version of this envelope for all \( b \in \{1, \ldots, B\} \).

One can think of \( c_\alpha(v) \), defined in (5), as truncating \( r(v) \) at an upper bound chosen to allow \( \hat{f}^* \) to exit from the “delete-bth-function” envelope for up to \( \lfloor B\alpha \rfloor \) values of \( b \).

### B.2 A More General Formulation

If (2) does not hold, then the above formulation may break down due to one or both of the following pathological situations:

1. For some \( v \), either \( \hat{f}^*_v(v) \leq \hat{f}(v) \leq \hat{f}^*_2(v) \) or \( \hat{f}^*_v(v) \leq \hat{f}(v) \leq \hat{f}^*_B(v) \) holds, where one of the pair of inequalities is strict. If, say, the first pair of inequalities holds, then \( \hat{f}^*_v(v) \) lies outside the pointwise interval
\[ \left[ \hat{f}(v) + K\{\hat{f}^*_v(v) - \hat{f}(v)\}, \hat{f}(v) + K\{\hat{f}^*_B(v) - \hat{f}(v)\} \right] \]
for any \( K > 0 \). In this sense, no amount of stretching is adequate at point \( v \), and accordingly it makes sense to set \( r(v) = \infty \) (see (8) below).

2. For some \( v \), \( \hat{f}^*_v(v), \ldots, \hat{f}^*_B(v) \) are all \( > \hat{f}(v) \) or all \( < \hat{f}(v) \). If this occurs, \( \hat{f}(v) \) would lie outside the stretched envelope (7).

These problems can be avoided by modifying the stretching procedure as follows. For each \( v \), let
\[ r(v) = \begin{cases} \infty, & \text{if } \hat{f}^*_v(v) \leq \hat{f}(v) \leq \hat{f}^*_2(v) \\
\max \left\{ \frac{\hat{f}^*_v(v) - \hat{f}(v)}{\hat{f}^*_v(v)}, \frac{\hat{f}^*_v(v) - \hat{f}(v)}{\hat{f}^*_2(v)}, \frac{\hat{f}^*_v(v) - \hat{f}(v)}{\hat{f}^*_B(v)} \right\}, & \text{otherwise.} \tag{8} \end{cases} \]
Similarly, for \( b = 1, \ldots, B \), let
\[
M_b = \begin{cases} 
\infty, & \text{if for some } v, \hat{f}_b^*(v) \leq \hat{f}(v) \leq \hat{f}_{(2)}(v) \\
\max_v \max \left\{ \frac{\hat{f}_v(v) - \hat{f}(v)}{\hat{f}_{(2)}(v) - \hat{f}(v)}, \frac{\hat{f}_v^*(v) - \hat{f}(v)}{\hat{f}_{(B-1)}(v) - \hat{f}(v)} \right\}, & \text{otherwise.}
\end{cases}
\]

We now define the stretched envelope as
\[
E_{\alpha}^c = \prod_v \left[ \hat{f}(v) + \min \left\{ 0, c_\alpha(v)[\hat{f}_1^*(v) - \hat{f}(v)] \right\}, \hat{f}(v) + \max \left\{ 0, c_\alpha(v)[\hat{f}_B^*(v) - \hat{f}(v)] \right\} \right],
\]
where \( c_\alpha(v) \) is given by (5) as before (and still satisfies (6)). This modification of (7) prevents \( \hat{f}(v) \) from lying outside the stretched envelope (the second pathological situation above). As with the simpler procedure above, stretching is necessary only when (3) obtains.

### B.3 Simultaneous Coverage of the Stretched Bands

As in the paper, define \( E_{\alpha}^c \) as the simultaneous interval constructed as in (9) using the same function \( c_\alpha \), but with \( \hat{f}_1^*(v) \) and \( \hat{f}_{B}(v) \) replaced by the minimum and maximum, respectively, of
\[
\{ \hat{f}_1^*(v), \ldots, \hat{f}_{B-1}(v), \hat{f}_{B+1}(v), \ldots, \hat{f}_B^*(v) \}.
\]
The following result says that, given (3) and an additional mild condition, (9) defines confidence bands with the desired \( 100(1 - \alpha)\% \) estimated simultaneous coverage.

**Proposition 2** Assume (3). If \( M([B(1-\alpha)]) < \infty \) then \( \hat{f}_b^* \) exists \( E_{\alpha}^c \) for at most \( B\alpha \) values \( b \in \{1, \ldots, B\} \).

**Proof.** Suppose \( \hat{f}_b^* \) exits \( E_{\alpha}^c \) at point \( v \). Without loss of generality \( \hat{f}_b^*(v) \) is above the upper limit of the envelope at \( v \). We claim that
\[
\hat{f}_b^*(v) = \hat{f}_{(B)}(v).
\]
Indeed, if this were not the case, the above would imply
\[
\hat{f}_b^*(v) > \hat{f}(v) + \max \left\{ 0, c_\alpha(v)[\hat{f}_{(B)}(v) - \hat{f}(v)] \right\}
\]
and thus
\[
0 < c_\alpha(v)[\hat{f}_{(B)}^*(v) - \hat{f}(v)] < \hat{f}^*_b(v) - \hat{f}(v) < \hat{f}_{(B)}^*(v) - \hat{f}(v),
\]
which contradicts (6). Thus (10) holds, and our initial supposition implies
\[
\hat{f}_b^*(v) = \hat{f}_{(B)}^*(v) > \hat{f}(v) + \max \left\{ 0, c_\alpha(v)[\hat{f}_{(B-1)}^*(v) - \hat{f}(v)] \right\}.
\]

(11)
If \( M_b < \infty \) then, by (8) and (11), \( r(v) \geq \frac{\hat{f}^{*}_{b}(v) - \hat{f}(v)}{\hat{f}^{*}_{(B-1)}(v) - \hat{f}(v)} > c_\alpha(v) \), whence, by (5), \( c_\alpha(v) = M_{\{B(1-\alpha)\}} \). It follows that
\[
M_b \geq \frac{\hat{f}^{*}_{b}(v) - \hat{f}(v)}{\hat{f}^{*}_{(B-1)}(v) - \hat{f}(v)} = \frac{\hat{f}^{*}_{(B)}(v) - \hat{f}(v)}{\hat{f}^{*}_{(B-1)}(v) - \hat{f}(v)} > c_\alpha(v) = M_{\{B(1-\alpha)\}}.
\]
Thus, whether or not (12) holds, we have \( M_b > M_{\{B(1-\alpha)\}} \). Since this inequality holds for at most \( B\alpha \) values \( b \in \{1, \ldots, B\} \), the proposition follows.

In summary, Proposition 2 demonstrates that, under the stated assumptions, (9) defines sensible 100(1 - \( \alpha \))% simultaneous confidence bands for \( f \).

Reference