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* $\boldsymbol{UDV}^T$ *be the singular value decomposition of the* $n \times K$ *matrix* $\boldsymbol{Z}$*, let* $\boldsymbol{U}_q$ *be the matrix consisting of the first* $q < \min(n, K)$ *columns of* $\boldsymbol{U}$*, and let* $\boldsymbol{D}_q$ *be the* $q \times q$ *upper left submatrix of* $\boldsymbol{D}$*. Then* $\boldsymbol{M}_0 = \boldsymbol{U}_q \boldsymbol{D}_q$ *minimizes*

$$\max_{\mathbf{w} \in \mathbf{R}^K, \|\mathbf{w}\|=1} \|\boldsymbol{Zw} - \mathit{proj}_{\mathbf{M}} \boldsymbol{Zw}\| \tag{1}$$

over all $n \times q$ *matrices* $\boldsymbol{M}$*, where* $\mathit{proj}_{\mathbf{M}}$ *denotes projection onto the column space of* $\boldsymbol{M}$*.*

Proof. Since $\mathrm{proj}_{\mathbf{M}} \boldsymbol{Zw} = \boldsymbol{M}(\boldsymbol{M}^T\boldsymbol{M})^{-1}\boldsymbol{M}^T\boldsymbol{Zw}$, (1) equals the spectral norm $\|\boldsymbol{Z} - \boldsymbol{M}(\boldsymbol{M}^T\boldsymbol{M})^{-1}\boldsymbol{M}^T\boldsymbol{Z}\|_2$. Extending the above notation, let $\boldsymbol{V}_q$ consist of the first q columns of $\boldsymbol{V}$, let $\boldsymbol{U}_{-q}$ consist of the last $K - q$ columns of $\boldsymbol{U}$, and similarly define $\boldsymbol{D}_{-q}$ and $\boldsymbol{V}_{-q}$. We then have $\boldsymbol{Z} = \boldsymbol{U}_q\boldsymbol{D}_q\boldsymbol{V}_q^T + \boldsymbol{U}_{-q}\boldsymbol{D}_{-q}\boldsymbol{V}_{-q}^T$, and thus

$$\begin{aligned} \mathrm{proj}_{\mathbf{M}_0} \boldsymbol{Zw} &= \boldsymbol{M}_0(\boldsymbol{M}_0^T\boldsymbol{M}_0)^{-1}\boldsymbol{M}_0^T\boldsymbol{Zw} \\ &= \boldsymbol{U}_q\boldsymbol{D}_q(\boldsymbol{D}_q\boldsymbol{U}_q^T\boldsymbol{U}_q\boldsymbol{D}_q)^{-1}\boldsymbol{D}_q\boldsymbol{U}_q^T \\ &\quad \times(\boldsymbol{U}_q\boldsymbol{D}_q\boldsymbol{V}_q^T + \boldsymbol{U}_{-q}\boldsymbol{D}_{-q}\boldsymbol{V}_{-q}^T)\boldsymbol{w}, \end{aligned}$$

which simplifies, using the equalities $\boldsymbol{U}_q^T\boldsymbol{U}_q = \boldsymbol{I}_q$ and $\boldsymbol{U}_q^T\boldsymbol{U}_{-q} = \mathbf{0}_{q\times(K-q)}$, to

$$\mathrm{proj}_{\mathbf{M}_0} \boldsymbol{Zw} = \boldsymbol{U}_q\boldsymbol{D}_q\boldsymbol{V}_q^T\boldsymbol{w}.$$

But by a standard result (e.g., Watkins, 2002), $\|\boldsymbol{Z} - \boldsymbol{U}_q \boldsymbol{D}_q \boldsymbol{V}_q^T\|_2 \le \|\boldsymbol{Z} - \boldsymbol{A}\|_2$ for any full-rank $n \times q$ matrix $\boldsymbol{A}$, and in particular for any such $\boldsymbol{A}$ of the form $\boldsymbol{M}(\boldsymbol{M}^T\boldsymbol{M})^{-1}\boldsymbol{M}^T\boldsymbol{Z}$ with $\boldsymbol{M}$ an $n \times q$ matrix. This establishes the claim of minimality. $\square$

B Stretched Confidence Bands

This supplementary appendix concerns simultaneous confidence bands formed by bootstrap estimates $\hat{\boldsymbol{f}}_1^*, \ldots, \hat{\boldsymbol{f}}_B^*$ 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

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

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 $v = v_b$

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

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 $\ge B(1-\alpha)$.) 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

$$c_\alpha(v) = \min\{r(v), M_{(\lceil B(1-\alpha)\rceil)}\}. \tag{5}$$

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

$$E^{c_\alpha} = \prod_v \left[\hat{f}(v) + c_\alpha(v)\{\hat{f}^*_{(1)}(v) - \hat{f}(v)\}, \hat{f}(v) + c_\alpha(v)\{\hat{f}^*_{(B)}(v) - \hat{f}(v)\}\right]. \tag{7}$$

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

$$\prod_v \left[\hat{f}(v) + r(v)\{\hat{f}^*_{(1)}(v) - \hat{f}(v)\}, \hat{f}(v) + r(v)\{\hat{f}^*_{(B)}(v) - \hat{f}(v)\}\right],$$

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 $\hat{\boldsymbol{f}}^*_b$ 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 $\hat{f}^*_{(1)}(v) \leq \hat{f}(v) \leq \hat{f}^*_{(2)}(v)$ or $\hat{f}^*_{(B-1)}(v) \leq \hat{f}(v) \leq \hat{f}^*_{(B)}(v)$ holds, where one of the pair of inequalities is strict. If, say, the first pair of inequalities holds, then $\hat{f}^*_{(1)}(v)$ lies outside the pointwise interval

   $$\left[\hat{f}(v) + K\{\hat{f}^*_{(2)}(v) - \hat{f}(v)\}, \hat{f}(v) + K\{\hat{f}^*_{(B)}(v) - \hat{f}(v)\}\right]$$

   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, $\hat{f}^*_{(1)}(v), \ldots, \hat{f}^*_{(B)}(v)$ 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

$$r(v) = \begin{cases} \infty, & \text{if } \hat{f}^*_{(1)}(v) \leq \hat{f}(v) \leq \hat{f}^*_{(2)}(v) \\ & \text{or } \hat{f}^*_{(B-1)}(v) \leq \hat{f}(v) \leq \hat{f}^*_{(B)}(v); \\ \max\left\{\frac{\hat{f}^*_{(1)}(v)-\hat{f}(v)}{\hat{f}^*_{(2)}(v)-\hat{f}(v)}, \frac{\hat{f}^*_{(B)}(v)-\hat{f}(v)}{\hat{f}^*_{(B-1)}(v)-\hat{f}(v)}\right\}, & \text{otherwise.} \end{cases} \tag{8}$$

Similarly, for $b = 1, \dots, B$, let

$$M_b = \begin{cases} \infty, & \text{if for some } v, \hat{f}_b^*(v) \le \hat{f}(v) \le \hat{f}_{(2)}^*(v) \\ & \text{or } \hat{f}_{(B-1)}^*(v) \le \hat{f}(v) \le \hat{f}_b^*(v); \\ \max_v \max\left\{ \frac{\hat{f}_b^*(v) - \hat{f}(v)}{\hat{f}_{(2)}^*(v) - \hat{f}(v)}, \frac{\hat{f}_b^*(v) - \hat{f}(v)}{\hat{f}_{(B-1)}^*(v) - \hat{f}(v)} \right\}, & \text{otherwise.} \end{cases}$$

We now define the stretched envelope as

$$E^{c_\alpha} = \prod_v \left[\hat{f}(v) + \min\left\{0, c_\alpha(v)[\hat{f}_{(1)}^*(v) - \hat{f}(v)]\right\}, \hat{f}(v) + \max\left\{0, c_\alpha(v)[\hat{f}_{(B)}^*(v) - \hat{f}(v)]\right\} \right], \tag{9}$$

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 $E_{-b}^{c_\alpha}$ as the simultaneous interval constructed as in (9) using the same function c_α, but with $\hat{f}_{(1)}^*(v)$ and $\hat{f}_{(B)}^*(v)$ replaced by the minimum and maximum, respectively, of

$$\{\hat{f}_1^*(v), \dots, \hat{f}_{b-1}^*(v), \hat{f}_{b+1}^*(v), \dots, \hat{f}_B^*(v)\}.$$

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 $M_{(\lceil B(1-\alpha)\rceil)} < \infty$ then $\hat{\boldsymbol{f}}_b^*$ exits $E_{-b}^{c_\alpha}$ for at most $B\alpha$ values $b \in \{1, \dots, B\}$.*

Proof. Suppose $\hat{\boldsymbol{f}}_b^*$ exits $E_{-b}^{c_\alpha}$ at point v. Without loss of generality $\hat{f}_b^*(v)$ is above the upper limit of the envelope at v. We claim that

$$\hat{f}_b^*(v) = \hat{f}_{(B)}^*(v). \tag{10}$$

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

$$\hat{f}_b^*(v) = \hat{f}_{(B)}^*(v) > \hat{f}(v) + \max\left\{0, c_\alpha(v)[\hat{f}_{(B-1)}^*(v) - \hat{f}(v)]\right\}. \tag{11}$$

If

$$M_b < \infty \tag{12}$$

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), $c_\alpha(v) = M_{(\lceil B(1-\alpha)\rceil)}$. It follows that

$$M_b \geq \frac{\hat{f}^*_b(v) - \hat{f}(v)}{\hat{f}^*_{(B-1)}(v) - \hat{f}(v)} = \frac{\hat{f}^*_{(B)}(v) - \hat{f}(v)}{\hat{f}^*_{(B-1)}(v) - \hat{f}(v)} > c_\alpha(v) = M_{(\lceil B(1-\alpha)\rceil)}.$$

Thus, whether or not (12) holds, we have $M_b > M_{(\lceil B(1-\alpha)\rceil)}$. Since this inequality holds for at most $B\alpha$ values $b \in \{1, \ldots, B\}$, the proposition follows. $\square$

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

Reference

Watkins, D. S. (2002). *Fundamentals of Matrix Computations*, 2nd ed. New York: John Wiley.