Intrinsic Regression Models for Positive-definite Matrices with Applications to Diffusion Tensor Imaging (Technical Details)

Abstract

In this technical report, we give detailed information about how to establish asymptotic theory for the intrinsic least-squares estimates of unknown parameters. We establish the consistency and asymptotic normality for the intrinsic least-squares estimates.

1 Assumptions

The following assumptions are needed to facilitate development of our methods, although they are not the weakest possible conditions.

(C1) $\boldsymbol{\beta}_*$ is an unique interior point of $\mathcal{B}$, where $\mathcal{B}$ is a compact set in R^p.

(C2) All functions in $\mathbf{C}_i(\boldsymbol{\beta})$ and $\Sigma_i(\boldsymbol{\beta})$ are three times continuously differentiable on $\mathcal{B}$. Let $g_i(\boldsymbol{\beta}) = \text{tr}(\mathcal{E}_i(\boldsymbol{\beta})^2)$, $||\partial_{\boldsymbol{\beta}} g_i(\boldsymbol{\beta})||^2$ and $||\partial^2_{\boldsymbol{\beta}} g_i(\boldsymbol{\beta})||$ are dominated by F_i, which is a function of $\mathbf{S}_i$ and $\mathbf{x}_i$. Moreover, $\mathbf{C}_i(\boldsymbol{\beta})$ and $\Sigma_i(\boldsymbol{\beta})$ are invertible matrices at least in an open neighborhood of $\boldsymbol{\beta}_*$.

(C3) For each $\epsilon > 0$, there exists a finite K such that $\sup_{n\geq 1} n^{-1} \sum_{i=1}^n E[F_i^2 \mathbf{1}(F_i > K)] < \epsilon$, where $\mathbf{1}(F_i > K)$ is the indicator function of $F_i > K$.

(C4) $\lim_{n\to\infty} n^{-1} E[-\partial^2_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta}_*)] = \mathbf{A}_1(\boldsymbol{\beta}_*)$ and $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n E\{[\partial_{\boldsymbol{\beta}} g_i(\boldsymbol{\beta}_*)]^{\otimes 2}\} = \mathbf{A}_2(\boldsymbol{\beta}_*)$, where $\mathbf{A}_1(\boldsymbol{\beta}_*)$ is positive definite.

Remarks 1: Assumptions (C1)-(C4) are standard conditions for ensuring the first order asymptotic properties (e.g., consistency and asymptotic normality) of M-estimators when the sample size is large (van der Vaart and Wellner 1996). With additional assumptions, we can establish higher order asymptotic results of saddlepoint approximations and the bootstrap approach, which are useful for small sample sizes. Due to space limitations, we will present them elsewhere.

Remarks 2: It should be noted that there may be multiple $\boldsymbol{\beta}_*$ that satisfy $\partial_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta}) = \mathbf{0}$ for some specifications of $\mathbf{C}(\mathbf{x}, \boldsymbol{\beta})$. For instance, for the first specification of $\mathbf{C}(\mathbf{x}, \boldsymbol{\beta})$ in $\partial_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta}) = \mathbf{0}$, $\boldsymbol{\beta}_1$ and $-\boldsymbol{\beta}_1$ are indistinguishable from the data. However, because these different true values $\boldsymbol{\beta}_*$ are isolated points in $\mathcal{B}$, we can establish asymptotic properties of $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}_*$, which are close to $\hat{\boldsymbol{\beta}}$ (Zhu and Zhang 2006). Numerically, the value of $\hat{\boldsymbol{\beta}}$ depends on the starting value of $\boldsymbol{\beta}$. In practice, we can also impose additional constraints to ensure uniqueness of $\boldsymbol{\beta}_*$ and $\hat{\boldsymbol{\beta}}$. For instance, for the first specification of $\mathbf{C}(\mathbf{x}, \boldsymbol{\beta})$ in $\partial_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta}) = \mathbf{0}$, we may standardize all the covariates and impose that the intercept term in $\mathbf{x}_i^T \boldsymbol{\beta}_1$ is greater than zero and thus the ambiguity of $\boldsymbol{\beta}_*$ and $\hat{\boldsymbol{\beta}}$ can be eliminated.

Remarks 3: Since $G_n(\boldsymbol{\beta})$ is generally not concave, it is difficult to guarantee that the estimator $\hat{\boldsymbol{\beta}}$ produced by our estimation procedure is the global minimum of $G_n(\boldsymbol{\beta})$. Similar to many statistical models including generalized linear models, we can only prove asymptotic existence and properties of a sequence of 'local minima' $\hat{\boldsymbol{\beta}}$ in a neighborhood of $\boldsymbol{\beta}_*$ (Gan and Jiang 1999; Fahrmeir and Kaufmann 1985; Small, Wang, and Yang 2003).

2 Proofs

First, we calculate the first and second derivatives of $G_n(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. We obtain Lemma 1 as follows.

Lemma 1. Under (C2), the following results hold:

(i) the a-th element of the $p \times 1$ vector $\partial_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta})$ is given by

$$\begin{aligned}\partial_{\beta_a} G_n(\boldsymbol{\beta}) &= -2\sum_{i=1}^{n} \mathrm{tr}\{\mathcal{E}_i(\boldsymbol{\beta})[\mathbf{C}_i(\boldsymbol{\beta})^{-1}\partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta}) + \partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta})^T\mathbf{C}_i(\boldsymbol{\beta})^{-T}]\} \qquad (1)\\ &= -2\sum_{i=1}^{n} \mathrm{tr}\{\mathcal{E}_i(\boldsymbol{\beta})\mathbf{C}_i(\boldsymbol{\beta})^{-1}\partial_{\beta_a}\Sigma_i(\boldsymbol{\beta})\mathbf{C}_i(\boldsymbol{\beta})^{-T}\},\end{aligned}$$

where $\partial_{\beta_a} = \partial/\partial\beta_a$, in which β_a is the a-th element of $\boldsymbol{\beta}$;

(ii) the (a, b)-th element of the $p \times p$ matrix $\partial_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta})$ is given by

$$\begin{aligned}\partial^2_{\beta_a\beta_b} G_n(\boldsymbol{\beta}) &= -2\sum_{i=1}^{n} \mathrm{tr}\{\partial_{\beta_b}\mathcal{E}_i(\boldsymbol{\beta})[\mathbf{C}_i(\boldsymbol{\beta})^{-1}\partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta}) + \partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta})^T\mathbf{C}_i(\boldsymbol{\beta})^{-T}]\}\\ &\quad -2\sum_{i=1}^{n} \mathrm{tr}\{\mathcal{E}_i(\boldsymbol{\beta})\partial_{\beta_b}[\mathbf{C}_i(\boldsymbol{\beta})^{-1}\partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta}) + \partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta})^T\mathbf{C}_i(\boldsymbol{\beta})^{-T}]\},\end{aligned}$$

where $\partial^2_{\beta_a\beta_b} = \partial^2/\partial\beta_a\partial\beta_b$ and $\partial_{\beta_b}\mathcal{E}_i(\boldsymbol{\beta}) = \int_0^1 \mathbf{h}(s, \boldsymbol{\beta})ds$, in which

$$\mathbf{h}(s, \boldsymbol{\beta}) = \{[\tilde{\mathbf{S}}_i(\boldsymbol{\beta}) - \mathbf{I}_3]s + \mathbf{I}_3\}^{-1}\partial_{\beta_b}\tilde{\mathbf{S}}_i(\boldsymbol{\beta})\{[\tilde{\mathbf{S}}_i(\boldsymbol{\beta}) - \mathbf{I}_3]s + \mathbf{I}_3\}^{-1} \qquad (2)$$

and $\tilde{\mathbf{S}}_i(\boldsymbol{\beta}) = \exp(\mathcal{E}_i(\boldsymbol{\beta})) = \mathbf{C}_i(\boldsymbol{\beta})^{-1}\mathbf{S}_i\mathbf{C}_i(\boldsymbol{\beta})^{-T}$;

(iii) $\partial_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta}_*)$ is unbiased, that is, $E[\partial_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta}_*)|\mathbf{x}_i] = \mathbf{0}$, and the conditional expectation of $\partial^2_{ab} G_n(\boldsymbol{\beta}_*)$ given $\mathbf{x}_i$ equals

$$\begin{aligned} E[\partial^2_{\beta_a\beta_b} G_n(\boldsymbol{\beta}_*)|\mathbf{x}_i] &= -2\sum_{i=1}^n \mathrm{tr}\{E[\partial_{\beta_b}\mathcal{E}_i(\boldsymbol{\beta}_*)|\mathbf{x}_i][\mathbf{C}_i(\boldsymbol{\beta}_*)^{-1}\partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta}_*) + \partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta}_*)^T\mathbf{C}_i(\boldsymbol{\beta}_*)^{-T}]\} \\ &= -2\sum_{i=1}^n \mathrm{tr}\{E[\partial_{\beta_b}\mathcal{E}_i(\boldsymbol{\beta}_*)|\mathbf{x}_i][\mathbf{C}_i(\boldsymbol{\beta}_*)^{-1}\partial_{\beta_a}\Sigma_i(\boldsymbol{\beta}_*)\mathbf{C}_i(\boldsymbol{\beta}_*)^{-T}]\}. \quad (3) \end{aligned}$$

Proof of Lemma 1. (i) It follows from Proposition 2.1 in Maher (2005) that

$$\begin{aligned} \partial_{\beta_a} G_n(\boldsymbol{\beta}) &= \sum_{i=1}^n \partial_{\beta_a}\mathrm{tr}\{[\log(\mathbf{C}_i(\boldsymbol{\beta})^{-1}\mathbf{S}_i\mathbf{C}_i(\boldsymbol{\beta})^{-T})]^2\} \\ &= 2\sum_{i=1}^n \mathrm{tr}[\mathcal{E}_i(\boldsymbol{\beta})\mathbf{C}_i(\boldsymbol{\beta})^T\mathbf{S}_i^{-1}\mathbf{C}_i(\boldsymbol{\beta})\partial_{\beta_a}(\mathbf{C}_i(\boldsymbol{\beta})^{-1}\mathbf{S}_i\mathbf{C}_i(\boldsymbol{\beta})^{-T})]. \end{aligned}$$

Because $\tilde{\mathbf{S}}_i(\boldsymbol{\beta})\mathcal{E}_i(\boldsymbol{\beta}) = \exp(\mathcal{E}_i(\boldsymbol{\beta}))\mathcal{E}_i(\boldsymbol{\beta}) = \mathcal{E}_i(\boldsymbol{\beta})\tilde{\mathbf{S}}_i(\boldsymbol{\beta})$ and

$$\partial_{\beta_a}(\mathbf{C}_i(\boldsymbol{\beta})^{-1}\mathbf{S}_i\mathbf{C}_i(\boldsymbol{\beta})^{-T}) = -\mathbf{C}_i(\boldsymbol{\beta})^{-1}\partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta})\tilde{\mathbf{S}}_i(\boldsymbol{\beta}) - \tilde{\mathbf{S}}_i(\boldsymbol{\beta})\partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta})^T\mathbf{C}_i(\boldsymbol{\beta})^{-T},$$

we have

$$\begin{aligned} \partial_{\beta_a} G_n(\boldsymbol{\beta}) &= -2\sum_{i=1}^n \mathrm{tr}[\mathcal{E}_i(\boldsymbol{\beta})\tilde{\mathbf{S}}_i(\boldsymbol{\beta})^{-1}\mathbf{C}_i(\boldsymbol{\beta})^{-1}\partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta})\tilde{\mathbf{S}}_i(\boldsymbol{\beta}) + \mathcal{E}_i(\boldsymbol{\beta})\tilde{\mathbf{S}}_i(\boldsymbol{\beta})^{-1}\tilde{\mathbf{S}}_i(\boldsymbol{\beta})\partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta})^T\mathbf{C}_i(\boldsymbol{\beta})^{-T}] \\ &= -2\sum_{i=1}^n \mathrm{tr}[\tilde{\mathbf{S}}_i(\boldsymbol{\beta})\mathcal{E}_i(\boldsymbol{\beta})\tilde{\mathbf{S}}_i(\boldsymbol{\beta})^{-1}\mathbf{C}_i(\boldsymbol{\beta})^{-1}\partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta}) + \mathcal{E}_i(\boldsymbol{\beta})\partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta})^T\mathbf{C}_i(\boldsymbol{\beta})^{-T}] \\ &= -2\sum_{i=1}^n \mathrm{tr}[\mathcal{E}_i(\boldsymbol{\beta})(\mathbf{C}_i(\boldsymbol{\beta})^{-1}\partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta}) + \partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta})^T\mathbf{C}_i(\boldsymbol{\beta})^{-T})]. \end{aligned}$$

This proves Lemma 1 (i). With some additional calculations, we can get Lemma 1 (ii). Moreover, a proof of (2) can be found in Dieci, Morini, and Papini (1996). Since $E[\mathcal{E}_i(\boldsymbol{\beta})|\mathbf{x}_i] = \mathbf{0}$, we obtain Lemma 1 (iii).

Lemma 1 gives the explicit forms of the first and second order derivatives of $G_n(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. In practice, we suggest approximating $\partial^2_{\beta_a\beta_b} G_n(\boldsymbol{\beta})$ by the first term $-2\sum_{i=1}^n \mathrm{tr}\{\partial_{\beta_b}\mathcal{E}_i(\boldsymbol{\beta})[\mathbf{C}_i(\boldsymbol{\beta})^{-1}\partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta}) + \partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta})^T\mathbf{C}_i(\boldsymbol{\beta})^{-T}]\}$, since the expectation of

$$-2\sum_{i=1}^n \mathrm{tr}\{\mathcal{E}_i(\boldsymbol{\beta})\partial_{\beta_b}[\mathbf{C}_i(\boldsymbol{\beta})^{-1}\partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta}) + \partial_{\beta_a}\mathbf{C}_i(\boldsymbol{\beta})^T\mathbf{C}_i(\boldsymbol{\beta})^{-T}]\}$$

equals zero at $\boldsymbol{\beta}_*$. In addition, computing $\partial^2_{\beta_a\beta_b} G_n(\boldsymbol{\beta})$ involves a one-dimensional integration of $\partial_{\beta_b}\mathcal{E}_i(\boldsymbol{\beta})$ in (2). Throughout the paper, we use the composite trapezoidal rule for 100 equally spaced nodes to approximate $\partial_{\beta_b}\mathcal{E}_i(\boldsymbol{\beta})$ (Burden and Faires 2000).

Proof of Theorem 1. We prove Theorem 1 in two parts. The first part proves weak consistency of $\hat{\boldsymbol{\beta}}$. We only need to show that

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}} n^{-1} |G_n(\boldsymbol{\beta}) - E[G_n(\boldsymbol{\beta})]| \to 0 \tag{4}$$

in probability and $E[G_n(\boldsymbol{\beta})]$ is continuous in $\boldsymbol{\beta}$ uniformly over $\boldsymbol{\beta} \in \Theta$. Conditions (C2) and (C3) are sufficient for Assumption W-LIP in Andrews (1992), which ensures the continuity of $E[G_n(\boldsymbol{\beta})]$ and the stochastic equicontinuity (SE) of $G_n(\boldsymbol{\beta})$. Furthermore, Conditions (C2) and (C3) ensure pointwise convergence, that is, $n^{-1}\{G_n(\boldsymbol{\beta}) - E[G_n(\boldsymbol{\beta})]\}$ converges to zero for each $\boldsymbol{\beta}$ in probability. Thus, combining SE and pointwise convergence yields (4). Combining (4) and (C1) ensures weak consistency of $\hat{\boldsymbol{\beta}}$.

The second part shows that

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*) = \mathbf{A}_1(\boldsymbol{\beta}_*)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \partial_{\boldsymbol{\beta}} g_i(\boldsymbol{\beta}_*) + o_p(1). \tag{5}$$

Conditions (C1)-(C4) are sufficient for establishing (5) (Andrews 1999). Applying the Lindeberg Theorem completes the proof of Theorem 1.

Proof of Theorem 2. To derive the score test statistic W_n, we need additional notation as follows. Without loss of generality, we assume that $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2)$, in which $\mathbf{R}_1$ is an $r \times r$ nonsingular matrix and $\mathbf{R}_2$ is an $r \times (p-r)$ matrix. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_{(1)}^T, \boldsymbol{\beta}_{(2)}^T)^T$, where $\boldsymbol{\beta}_{(1)}$ is an $r \times 1$ vector corresponding to $\mathbf{R}_1$ and $\boldsymbol{\beta}_{(2)}$ is a $(p-r) \times 1$ vector corresponding to $\mathbf{R}_2$. If we define $\boldsymbol{\mu} = \mathbf{R}_1 \boldsymbol{\beta}_{(1)} + \mathbf{R}_2 \boldsymbol{\beta}_{(2)} - \boldsymbol{b}_0$, then there exists a one-to-one correspondence between $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\beta}_{(2)}) = \mathbf{f}(\boldsymbol{\beta})$ and $\boldsymbol{\beta} = \mathbf{f}^{-1}(\boldsymbol{\theta})$. Thus, we have

$$\frac{\partial(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial(\boldsymbol{\mu}, \boldsymbol{\beta}_2)} = \begin{pmatrix} \mathbf{R}_1^{-1} & -\mathbf{R}_1^{-1}\mathbf{R}_2 \\ \mathbf{0} & \mathbf{I}_{p-r} \end{pmatrix}.$$

Moreover, the first- and second-order derivatives of $G_n(\boldsymbol{\beta})$ with respect to $\boldsymbol{\mu}$ are given by $\partial_{\boldsymbol{\mu}} G_n(\boldsymbol{\beta}) = (\mathbf{R}_1^{-1}, \mathbf{0}) \partial_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta})$,

$$\partial^2_{\boldsymbol{\mu}\boldsymbol{\beta}} G_n(\boldsymbol{\beta}) = (\mathbf{R}_1^{-1}, \mathbf{0}) \partial^2_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta}),$$

and

$$\partial^2_{\boldsymbol{\mu}} G_n(\boldsymbol{\beta}) = (\mathbf{R}_1^{-1}, \mathbf{0})\partial^2_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta})(\mathbf{R}_1^{-1}, \mathbf{0})^T.$$

We obtain the asymptotic distribution of the test statistic W_n as follows. Let $\boldsymbol{\theta}_* = (\mathbf{0}, \boldsymbol{\beta}_{(2)*})$ be the true parameter vector of $\boldsymbol{\beta}$ under H_0 and $\tilde{\boldsymbol{\theta}} = \mathbf{f}(\mathbf{0}, \tilde{\boldsymbol{\beta}}_{(2)})$ is the ILSE of $\boldsymbol{\theta}$ under H_0. Assume that $G_n(\boldsymbol{\theta}) = G_n(\mathbf{f}^{-1}(\boldsymbol{\theta}))$ and

$$-\partial^2_{\boldsymbol{\theta}} G_n(\boldsymbol{\theta}) = \mathbf{V}(\boldsymbol{\mu}, \boldsymbol{\beta}_{(2)}) = \begin{pmatrix} \mathbf{V}_{\boldsymbol{\mu}\boldsymbol{\mu}} & \mathbf{V}_{\boldsymbol{\mu}\boldsymbol{\beta}_{(2)}} \\ \mathbf{V}_{\boldsymbol{\beta}_{(2)}\boldsymbol{\mu}} & \mathbf{V}_{\boldsymbol{\beta}_{(2)}\boldsymbol{\beta}_{(2)}} \end{pmatrix}.$$

Similar to Theorem 1, we have $\tilde{\boldsymbol{\beta}}_{(2)} - \boldsymbol{\beta}_{(2)*} = \mathbf{V}^{-1}_{\boldsymbol{\beta}_{(2)}\boldsymbol{\beta}_{(2)}} \partial_{\boldsymbol{\beta}_{(2)}} G_n(\boldsymbol{\theta}_*)[1 + o_p(1)]$. Second, using a Taylor's series expansion leads to

$$\partial_{\boldsymbol{\mu}} G_n(\tilde{\boldsymbol{\theta}}) = \partial_{\boldsymbol{\mu}} G_n(\boldsymbol{\theta}_*) - \mathbf{V}_{\boldsymbol{\mu}\boldsymbol{\beta}_{(2)}} \mathbf{V}^{-1}_{\boldsymbol{\beta}_{(2)}\boldsymbol{\beta}_{(2)}} \partial_{\boldsymbol{\beta}_{(2)}} G_n(\boldsymbol{\theta}_*)[1 + o_p(1)] = \sum_{i=1}^n \mathbf{U}_i(\mathbf{0}, \tilde{\boldsymbol{\beta}}_{(2)})[1 + o_p(1)],$$

where $\mathbf{U}_i(\boldsymbol{\theta}) = \partial_{\boldsymbol{\mu}} g_i(\mathbf{f}^{-1}(\boldsymbol{\theta})) - \mathbf{V}_{\boldsymbol{\mu}\boldsymbol{\beta}_{(2)}} \mathbf{V}^{-1}_{\boldsymbol{\beta}_{(2)}\boldsymbol{\beta}_{(2)}} \partial_{\boldsymbol{\beta}_{(2)}} g_i(\mathbf{f}^{-1}(\theta))$. We define $\hat{\mathbf{U}}_{i,\boldsymbol{\mu}}(\tilde{\boldsymbol{\beta}}) = \partial_{\boldsymbol{\mu}} g_i(\tilde{\boldsymbol{\beta}}) - \hat{\mathbf{V}}_{\boldsymbol{\mu}\boldsymbol{\beta}_{(2)}} \hat{\mathbf{V}}^{-1}_{\boldsymbol{\beta}_{(2)}\boldsymbol{\beta}_{(2)}} \partial_{\boldsymbol{\beta}_{(2)}} g_i(\tilde{\boldsymbol{\beta}})$, where $\hat{\mathbf{V}}_{\boldsymbol{\mu}\boldsymbol{\beta}_{(2)}} = \mathbf{V}_{\boldsymbol{\mu}\boldsymbol{\beta}_{(2)}}(\mathbf{0}, \tilde{\boldsymbol{\beta}}_{(2)})$ and $\mathbf{V}_{\boldsymbol{\beta}_{(2)}\boldsymbol{\beta}_{(2)}} = \mathbf{V}_{\boldsymbol{\beta}_{(2)}\boldsymbol{\beta}_{(2)}}(\mathbf{0}, \tilde{\boldsymbol{\beta}}_{(2)})$. We define $\hat{\mathbf{I}}_{\boldsymbol{\mu}\boldsymbol{\mu}} = n^{-1} \sum_{i=1}^n [\hat{\mathbf{U}}_{i,\boldsymbol{\mu}}(\tilde{\boldsymbol{\beta}}) - \overline{\mathbf{U}}_{\boldsymbol{\mu}}(\tilde{\boldsymbol{\beta}})][\hat{\mathbf{U}}_{i,\boldsymbol{\mu}}(\tilde{\boldsymbol{\beta}}) - \overline{\mathbf{U}}_{\boldsymbol{\mu}}(\tilde{\boldsymbol{\beta}})]^T$, where $\overline{\mathbf{U}}_{\boldsymbol{\mu}}(\tilde{\boldsymbol{\beta}}) = \sum_{i=1}^n \hat{\mathbf{U}}_{i,\boldsymbol{\mu}}(\tilde{\boldsymbol{\beta}})/n$. Under (C1)-(C4), $n^{-1/2} \hat{\mathbf{I}}^{-1/2}_{\boldsymbol{\mu}\boldsymbol{\mu}} \partial_{\boldsymbol{\mu}} G_n(\tilde{\theta})$ converges to a Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{I}_r$; consequently, W_n is asymptotically distributed as a $\chi^2(r)$ distribution under H_0.

3 Figure for Significance Testing of Age Effect

REFERENCES

Andrews, D. W. K. (1992), "Generic Uniform Convergence," *Econometric Theory*, 8, 241-57.

Andrews, D. W. K. (1999), "Estimation When a Parameter is on a Boundary: Theory and Applications," *Econometrica*, 67, 1341-1383.

Burden, R. L., and Faires, J. D. (2000), *Numerical Analysis,* 7th Ed., Brooks/Cole.

Dieci, L., Morini, B., and Papini, A. (1996), "Computational Techniques for Real Logarithms of Matrices," *SIAM J. Matrix Anal. Appl.*, 17, 570-593.

Do Carmo, M. P. (1992), *Riemannian Geometry.* Boston, Birkhauser.

Fahrmeir, L., and Kaufmann, H. (1985), "Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models," *Annals of Statistics*, 13, 342-368.

Gan, L. and Jiang, J. (1999), "A Test of Global Maximum," *Journal of the American Statistical Association*, 94, 847-854.

Maher, M. (2005), "A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-definite Matrices," *SIAM Journal of Matrix Analysis and Applications*, 26, 735-747.

Small, C.G., Wang, J. and Yang, Z. (2000), "Eliminating Multiple Root Problems in Estimation (with Discussions)," *Statistical Science*, 15, 313-341.

van der Vaart and Wellner J. (1996), *Weak Convergence and Empirical Processes.* New York: Springer.

Zhu, H. T., and Zhang, H. P. (2006), "Asymptotics for Estimation and Testing Under Loss of Identifiability," *Journal of Multivariate Analysis*, 97, 19-45.

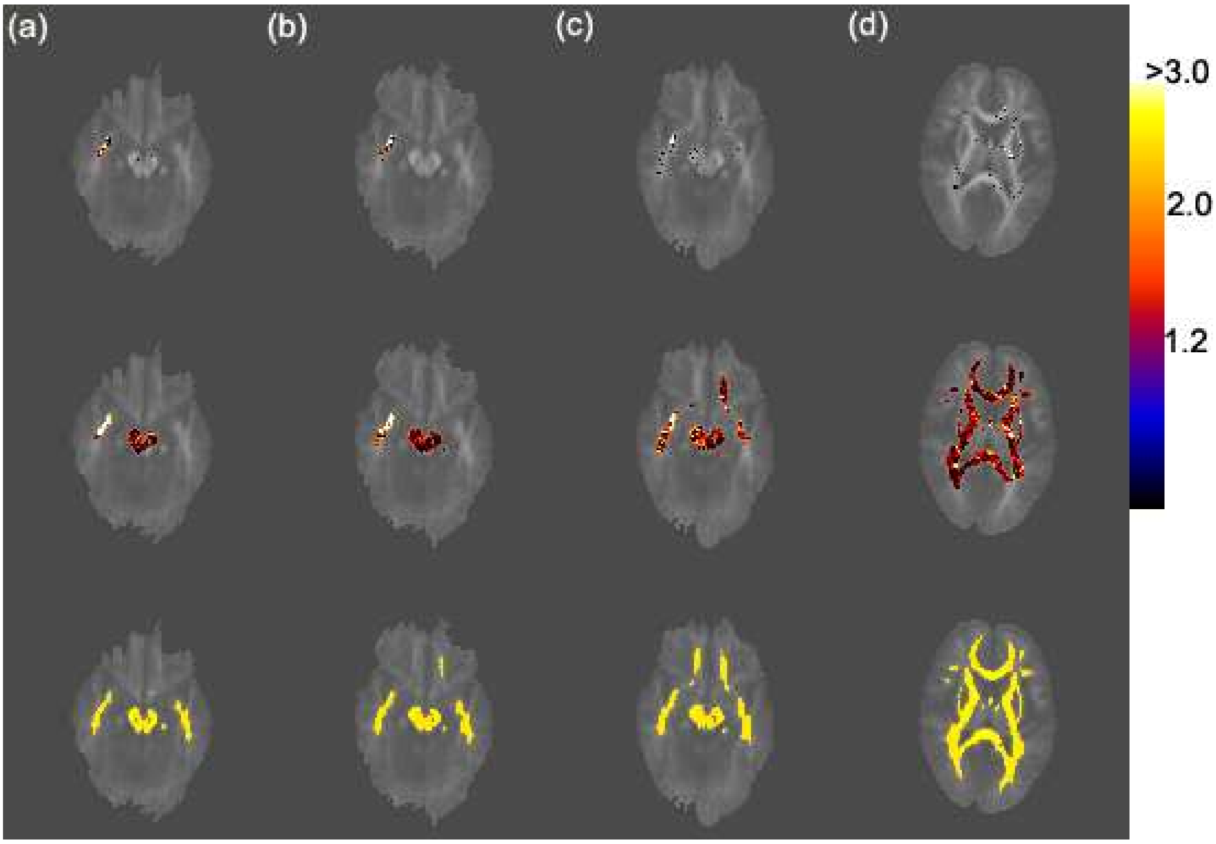

Figure 1: Significance testing of age effect: color-coded maps of raw and adjusted p-values in four selected ROIs of the reference brain. The color scale reflects the magnitude of the values of $-\log_{10}(P)$, with black to blue representing smaller values (0-1) and red to white representing larger values (1.88-3). Row 1: adjusted $-\log_{10}(P)$ values of the score statistics based on our test procedure for the correction of multiple comparisons. Row 2: raw $-\log_{10}(P)$ values of the score statistics based on a χ^2 distribution. Row 3: selected ROIs with FA values greater than 0.4. After correcting for multiple comparisons, statistically significant age effects remain in the inferior longitudinal fasciculus.