

A SHORT GUIDE TO
SEMIPARAMETRIC INFERENCE
TSIATIS CHAPTERS 3 AND 4

Consider the Hilbert space $\mathcal{H}$ of q -dimensional functions of random variable Z , with

$$\mathbb{E}[h(Z)] = 0_{q \times 1} \quad \mathbb{E}[\{h(Z)\}^\top h(Z)] < \infty$$

for $h \in \mathcal{H}$, with inner product

$$\langle h_1, h_2 \rangle = \mathbb{E}[\{h_1(Z)\}^\top h_2(Z)] \quad h_1, h_2 \in \mathcal{H}.$$

Let $\|h\|^2 = \langle h, h \rangle$. We say that h_1 and h_2 are *orthogonal* if

$$\langle h_1, h_2 \rangle = 0.$$

For linearly independent functions $h_1, \dots, h_k \in \mathcal{H}$ consider the linear subspace, $\mathcal{U}$ of $\mathcal{H}$ defined by

$$\mathcal{U} \equiv \left\{ u : u = \sum_{j=1}^k a_j h_j, (a_1, \dots, a_k) \in \mathbb{R}^k \right\}.$$

For each $h \in \mathcal{H}$, there is a unique value $u_0 \in \mathcal{U}$ such that

$$\|h - u_0\| < \|h - u\| \quad \forall u \in \mathcal{U}, u \neq u_0.$$

- u_0 is the *projection* of h onto $\mathcal{U}$ (denoted $\Pi(h|\mathcal{U})$);
- $h - u_0$ is *orthogonal* to $\mathcal{U}$;
- $\|h\|^2 = \|u_0\|^2 + \|h - u_0\|^2$.

Suppose that $v \equiv v(Z)$ is a r -dimensional function of Z with

$$\mathbb{E}[v(Z)] = 0_{r \times 1} \quad \mathbb{E}[\{v(Z)\}^\top v(Z)] < \infty.$$

Let

$$\mathcal{U}_v \equiv \{Bv : B \text{ an arbitrary } q \times r \text{ matrix}\}$$

For $h \in \mathcal{H}$, the projection of h onto $\mathcal{U}_v$ is given by B_0v satisfying

$$\langle h - B_0v, Bv \rangle = 0 \quad \forall B \in \mathbb{R}^{q \times r}$$

that is,

$$B_0 = \mathbb{E}[hv^\top] \left\{ \mathbb{E}[vv^\top] \right\}^{-1} \quad B_0v = \mathbb{E}[hv^\top] \left\{ \mathbb{E}[vv^\top] \right\}^{-1} v.$$

Parametric Inference: Set up

Suppose $Z_1, \dots, Z_n$ are i.i.d. with density $p_Z(z; \theta)$ with $\theta \in \mathbb{R}^p$, and

$$\theta = \begin{bmatrix} \beta \\ \eta \end{bmatrix} \quad \begin{array}{l} q \times 1 \\ r \times 1 \end{array}$$

with true value θ_0 comprised of β_0 and η_0 .

- β – parameter of interest;
- η – nuisance parameter.

Step 1

Asymptotically Linear Estimator: An estimator $\hat{\beta}_n$ of parameter β_0 is *asymptotically linear* if there exists a q -dimensional function, $\varphi(Z)$, with

$$\mathbb{E}_{\beta_0}[\varphi(Z)] = 0_{q \times 1} \quad \mathbb{E}_{\beta_0}[\varphi(Z)\{\varphi(Z)\}^\top] < \infty, \text{ nonsingular}$$

such that

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(Z_i) + o_p(1).$$

Then as

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(Z_i) \xrightarrow{d} \text{Normal} \left(0_{q \times 1}, \mathbb{E}_{\beta_0}[\varphi(Z)\{\varphi(Z)\}^\top] \right)$$

$\sqrt{n}(\hat{\beta}_n - \beta_0)$ also has this asymptotic distribution. If such a $\varphi(\cdot)$ exists, it is unique; it is termed the *influence function*.

Step 2

Regular Estimator: Consider the *local data generating process* (LDGP)

$$Z_{1n}, \dots, Z_{nn} \sim p_Z(z; \theta_n) \quad \text{i.i.d.}$$

with $\sqrt{n}(\theta_n - \theta^*) \rightarrow \text{constant}$ as $n \rightarrow \infty$. Estimator $\hat{\beta}_n$ of β_n is *regular* if its limiting distribution does not depend on θ_n . That is, in the Normal case, $\hat{\beta}_n$ is regular if

$$\sqrt{n}(\hat{\beta}_n - \beta^*) \xrightarrow{d} \text{Normal}(0, \Sigma^*)$$

under $p_Z(z; \theta^*)$ implies that

$$\sqrt{n}(\hat{\beta}_n - \beta_n) \xrightarrow{d} \text{Normal}(0, \Sigma^*)$$

under $p_Z(z; \theta_n)$.

Score Function: The *score function* $S_{\theta}(z, \theta_0)$ is defined by

$$S_{\theta}(z, \theta_0) = \frac{\partial}{\partial \theta} \{\log p_Z(z; \theta)\}_{\theta=\theta_0} = \begin{bmatrix} S_{\beta}(z, \theta_0) \\ S_{\eta}(z, \theta_0) \end{bmatrix} \quad \begin{matrix} q \times 1 \\ r \times 1 \end{matrix}$$

with the subscript denoting the variable with respect to which the derivative is being taken.

Theorem: If $\beta(\theta)$ is a q -dimensional parameter of interest, let

$$\Gamma(\theta) = \frac{\partial \beta(\theta)}{\partial \theta^\top} \quad q \times p.$$

Suppose that $\hat{\beta}_n$ is regular and asymptotically linear (RAL) with influence function $\varphi(\cdot)$ such that $\mathbb{E}_{\theta_0}[\{\varphi(Z)\}^\top \varphi(Z)] < \infty$. Then

$$\mathbb{E}_{\theta_0}[\varphi(Z) \{S_\theta(Z, \theta_0)\}^\top] = \Gamma(\theta_0).$$

Step 4 (cont.)

If $\theta = (\beta^\top, \eta^\top)^\top$, then

$$\Gamma(\theta_0) = \begin{bmatrix} \mathbf{I}_{q \times q} & 0_{q \times r} \\ 0_{r \times q} & 0_{r \times r} \end{bmatrix}$$

and so

(i) $\mathbb{E}_{\theta_0}[\varphi(Z) \{S_\beta(z, \theta_0)\}^\top] = \mathbf{I}_{q \times q};$

(ii) $\mathbb{E}_{\theta_0}[\varphi(Z) \{S_\eta(z, \theta_0)\}^\top] = 0_{q \times r}.$

That is, $\varphi(Z)$ is **orthogonal** to $S_\eta(z, \theta_0)$.

Step 5

***m*-estimation:** Recall that in likelihood estimation, we have the log-likelihood

$$\ell_n(\theta) = \sum_{i=1}^n \log p_Z(z_i; \theta).$$

Under regularity conditions, by the mean-value theorem, we have

$$\dot{\ell}_n(\theta) = \dot{\ell}_n(\theta_0) + \ddot{\ell}_n(\theta')(\theta - \theta_0)$$

where $\|\theta' - \theta_0\| < \|\theta - \theta_0\|$, and where

$$\dot{\ell}_n(\theta) = \frac{\partial \ell_n(\theta)}{\partial \theta} \qquad \ddot{\ell}_n(\theta) = \frac{\partial^2 \ell_n(\theta)}{\partial \theta \partial \theta^\top}.$$

Step 5 (cont.)

Evaluating at $\theta = \hat{\theta}_n$, and noting that $\dot{\ell}_n(\hat{\theta}_n) = 0_{p \times 1}$, we have on rearrangement and multiplying through by $1/\sqrt{n}$ that

$$\left\{ -\frac{1}{n} \ddot{\ell}_n(\theta) \right\} \sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \dot{\ell}_n(\theta_0).$$

As $n \rightarrow \infty$, we have that

$$\left\{ \frac{1}{n} \ddot{\ell}_n(\theta) \right\} \xrightarrow{p} \mathbb{E}_{\theta_0} \left[\frac{\partial^2 \log p_Z(z; \theta)}{\partial \theta \partial \theta^\top} \Big|_{\theta = \theta_0} \right] = \mathbb{E}_{\theta_0} [\ddot{\ell}(\theta_0)]$$

say.

Therefore

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n - \theta_0) &= \left\{ -\mathbb{E}_{\theta_0} \left[\ddot{\ell}(\theta_0) \right] \right\}^{-1} \frac{1}{\sqrt{n}} \dot{\ell}_n(\theta_0) + \mathbf{o}_p(1) \\ &= \left\{ -\mathbb{E}_{\theta_0} \left[\ddot{\ell}(\theta_0) \right] \right\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n S_{\theta}(Z_i, \theta_0) + \mathbf{o}_p(1)\end{aligned}$$

which yields the influence function

$$\varphi(Z) = \left\{ -\mathbb{E}_{\theta_0} \left[\ddot{\ell}(\theta_0) \right] \right\}^{-1} S_{\theta}(Z, \theta_0).$$

Step 5 (cont.)

In m -estimation, we replace the *score equation*

$$\sum_{i=1}^n S_{\theta}(Z_i, \theta) = 0_{p \times 1}$$

by the more general form

$$\sum_{i=1}^n m(Z_i, \theta) = 0_{p \times 1}$$

for function $m(., .)$ with the properties $\mathbb{E}_{\theta}[m(Z, \theta)] = 0_{p \times 1}$ and

$$\mathbb{E}_{\theta}[\{m(Z, \theta)\}^{\top} m(Z, \theta)] < \infty \quad \mathbb{E}_{\theta}[m(Z, \theta)\{m(Z, \theta)\}^{\top}] \text{ nonsingular}$$

for all θ .

Then using exactly the same logic as in the likelihood case, we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(Z_i) + o_p(1)$$

for influence function

$$\varphi(Z) = \{-\mathbb{E}_{\theta_0} [\dot{m}(Z, \theta_0)]\}^{-1} m(Z, \theta_0).$$

Step 5 (cont.)

We have by the CLT and elementary results for the Normal distribution that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \longrightarrow \text{Normal}(0_{p \times 1}, B^{-1}AB^{-1})$$

where

$$A \equiv A(\theta_0) = \mathbb{E}_{\theta_0}[m(Z, \theta_0)\{m(Z, \theta_0)\}^\top]$$

and

$$B \equiv B(\theta_0) = \mathbb{E}_{\theta_0}[\dot{m}(Z, \theta_0)].$$

These $p \times p$ matrices are typically estimated by

$$\hat{A}_n = \frac{1}{n} \sum_{i=1}^n m(Z_i, \theta_0)\{m(Z_i, \theta_0)\}^\top \quad \hat{B}_n = \frac{1}{n} \sum_{i=1}^n \dot{m}(Z_i, \theta_0).$$

The Geometry of Influence Functions: Consider

- $\mathcal{H}$: usual Hilbert space of zero-mean, finite variance q -dimensional functions with the usual inner product.
- $\mathcal{T}$: linear subspace $\mathcal{T}$ of $\mathcal{H}$ – the *tangent space* for the probability model $p_Z(z; \theta)$, defined by

$$\mathcal{T} \equiv \{BS_\theta(z, \theta) : B \text{ a } q \times p \text{ matrix}\}$$

- Λ : linear subspace Λ of $\mathcal{H}$ – the *nuisance tangent space* for the probability model $p_Z(z; \theta)$, defined by

$$\Lambda \equiv \{BS_\eta(z, \theta) : B \text{ a } q \times r \text{ matrix}\}$$

Step 6 (cont.)

By the previous theorem, if $\widehat{\beta}_n$ is an RAL estimator with influence function $\varphi_{\widehat{\beta}_n}(Z)$, then we must have

$$\mathbb{E}_{\theta_0} \left[\varphi_{\widehat{\beta}_n}(Z) \{S_\eta(Z, \theta)\}^\top \right] = 0_{q \times r}$$

that is, $\varphi_{\widehat{\beta}_n}(Z)$ is orthogonal to the nuisance tangent space. We write

$$\varphi_{\widehat{\beta}_n}(Z) \in \Lambda^\perp.$$

Step 6 (cont.)

We have that

$$\mathcal{H} = \Lambda \oplus \Lambda^\perp$$

that is, $h \in \mathcal{H}$ can be decomposed

$$h = a_1 + a_2 \quad a_1 \in \Lambda, a_2 \in \Lambda^\perp$$

with a_1 orthogonal to a_2 .

For $h \in \mathcal{H}$, write a_0 for the projection of h onto Λ

$$\Pi(h|\Lambda) = a_0$$

then as $h = a_0 + (h - a_0)$, by the projection theorem we have that

$$\Pi(h|\Lambda^\perp) = h - a_0.$$

By the result in the Preliminaries

$$\Pi(h|\Lambda) = \mathbb{E}[hS_\eta^\top] \left\{ \mathbb{E}[S_\eta S_\eta^\top] \right\}^{-1} S_\eta$$

so therefore all elements of $\Lambda^\perp$ take the form

$$h - \Pi(h|\Lambda) = h - \mathbb{E}[hS_\eta^\top] \left\{ \mathbb{E}[S_\eta S_\eta^\top] \right\}^{-1} S_\eta.$$

Finally we have

$$\mathcal{T} = \mathcal{T}_\beta \oplus \Lambda$$

where

$$\mathcal{T}_\beta \equiv \{BS_\beta(Z, \theta_0) : B \text{ a } q \times q \text{ matrix}\}.$$

Constructing an Estimating Function: Suppose influence function $\varphi(Z)$ satisfies the earlier Theorem. Set

$$m(Z, \beta, \eta) = \varphi(Z) - \mathbb{E}_{\beta, \eta}[\varphi(Z)]$$

which has mean zero, finite variance. Then $\hat{\beta}_n$ satisfying

$$\sum_{i=1}^n m(Z_i, \hat{\beta}_n, \hat{\eta}_n(\hat{\beta}_n)) = 0_{p \times 1}$$

is RAL with influence function $\varphi(Z)$.

Efficient Influence Function: The *efficient influence function*, $\varphi_{\text{eff}}(Z)$, is the influence function with the smallest variance. For an arbitrary influence function $\varphi^*(Z)$, we have that

$$\varphi_{\text{eff}}(Z) = \varphi^*(Z) - \Pi(\varphi^*(Z)|\mathcal{T}^\perp) \equiv \Pi(\varphi^*(Z)|\mathcal{T}).$$

In fact, if we denote

$$I(\theta_0) = \mathbb{E}_{\theta_0} \left[S_\theta(Z, \theta_0) \{S_\theta(Z, \theta_0)\}^\top \right]$$

we have explicitly that

$$\varphi_{\text{eff}}(Z) = \Gamma(\theta_0) \{I(\theta_0)\}^{-1} S_\theta(Z, \theta_0)$$

Step 8 (cont.)

The *efficient score function* for β is obtained by projecting $S_\beta(Z, \theta_0)$ onto the nuisance tangent space Λ , and taking the residual. Recall by the Preliminary result (p. 3)

$$\Pi(S_\beta(Z, \theta_0)|\Lambda) = \mathbb{E}[S_\beta S_\eta^\top] \left\{ \mathbb{E}[S_\eta S_\eta^\top] \right\}^{-1} S_\eta(Z, \theta_0)$$

so therefore the efficient score function for β is

$$S_{\text{eff}}(Z, \theta_0) = S_\beta(Z, \theta_0) - \mathbb{E}[S_\beta S_\eta^\top] \left\{ \mathbb{E}[S_\eta S_\eta^\top] \right\}^{-1} S_\eta(Z, \theta_0)$$

The efficient influence function for β is

$$\varphi_{\text{eff}}(Z) = \left\{ \mathbb{E}[S_{\text{eff}} S_{\text{eff}}^{\top}] \right\}^{-1} S_{\text{eff}}(Z, \theta_0)$$

which has variance

$$\left\{ \mathbb{E}[S_{\text{eff}} S_{\text{eff}}^{\top}] \right\}^{-1}$$

This result is the generalization of the results on p. 12 and p. 14.

Semiparametric case: Consider the class of semiparametric models

$$\mathcal{P} \equiv \{p_Z(z, \beta, \eta(.)) : \beta \text{ is } q \times 1, \eta(.) \text{ is infinite dimensional}\}$$

with true model $p_0(z) = p_Z(z, \beta_0, \eta_0(.))$.

Consider the *parametric submodel*

$$\mathcal{P}_{\beta, \gamma} \equiv \{p_Z(z, \beta, \gamma) : \beta \text{ is } q \times 1, \gamma \text{ is } r \times 1\}$$

where

- (i) $\mathcal{P}_{\beta, \gamma} \subset \mathcal{P}$;
- (ii) $p_0(z) \in \mathcal{P}_{\beta, \gamma}$ (the parametric submodel contains the truth). That is

$$p_0(z) \equiv p_Z(z, \beta_0, \gamma_0).$$

For example, for the *restricted moment model*

$$Y_i = \mu(X_i, \beta) + \epsilon_i$$

where $\eta(\cdot)$ specifies the density of ϵ_i admits the parametric submodel where

$$\epsilon_i \sim \text{Normal}(0, \sigma^2)$$

that is, $\gamma \equiv \sigma^2$.

Step 9 (cont.)

In the *proportional hazards model* with hazard function $\lambda(t|X) = \lambda(t) \exp\{\beta^\top X\}$ with $\lambda(t)$ nonparametrically specified, with true values of the parameters β_0 and $\lambda_0(t)$.

A parametric submodel takes the form

$$\lambda(t|X) = \lambda_0(t) \exp\{\gamma_1 h_1(t) + \cdots + \gamma_r h_r(t)\} \exp\{\beta^\top X\}$$

for specified functions $h_1(t), \dots, h_r(t)$.

This parametric model is specified in terms of $\lambda_0(t)$ which is not known.

However, if we specify $\beta = \beta_0$, and $\gamma_1 = \cdots = \gamma_r = 0$, we recover the true model.

Semiparametric inference: Using the same Hilbert space $\mathcal{H}$ of zero-mean q -dimensional functions with the covariance inner product. Define

(i) *parametric submodel nuisance tangent space*

$$\Lambda_\gamma \equiv \{BS_\gamma(Z, \beta_0, \gamma_0) : B \text{ a } q \times r \text{ matrix}\}$$

with $S_\gamma(Z, \beta_0, \gamma_0)$ being the score function component corresponding to γ from the parametric submodel.

(ii) *efficient influence function* for the parametric submodel

$$\varphi_{\beta,\gamma}^{\text{eff}}(Z) = \left\{ \mathbb{E} \left[S_{\beta,\gamma}^{\text{eff}}(Z, \beta_0, \gamma_0) \left\{ S_{\beta,\gamma}^{\text{eff}}(Z, \beta_0, \gamma_0) \right\}^{\top} \right] \right\}^{-1} S_{\beta,\gamma}^{\text{eff}}(Z, \beta_0, \gamma_0)$$

where the parametric submodel *efficient score function* is

$$S_{\beta,\gamma}^{\text{eff}}(Z, \beta_0, \gamma_0) = S_{\beta}(Z, \beta_0, \eta_0) - \Pi(S_{\beta}(Z, \beta_0, \eta_0) | \Lambda_{\gamma}).$$

Note that the RHS is defined in terms of (β_0, η_0) .

- (iii) The smallest asymptotic variance amongst RAL estimators for β is

$$\left\{ \mathbb{E} \left[S_{\beta, \gamma}^{\text{eff}}(Z, \beta_0, \gamma_0) \left\{ S_{\beta, \gamma}^{\text{eff}}(Z, \beta_0, \gamma_0) \right\}^{\top} \right] \right\}^{-1}$$

All expectations above are taken with respect to the true model

$$p_Z(z, \beta_0, \eta_0).$$

Step 10 (cont.)

- An estimator for β is RAL for the semiparametric model if it is RAL for every parametric submodel;
- Any influence function of an RAL estimator in the semiparametric model must be an influence function of an RAL estimator within a parametric submodel;
- Any influence function of an RAL estimator in the semiparametric submodel must be orthogonal to all parametric submodel nuisance tangent spaces;
- The variance of any RAL semiparametric influence function must be no smaller than the variance in (iii) above.

In the restricted mean model with Y a $d \times 1$ vector

$$Y = \mu(X, \beta) + \epsilon$$

we have

- $\mathcal{P} \equiv \{p_Z(z, \beta, \eta_0()), z = (y, x)\}$ with

$$p_{Y,Z}(y, x) \equiv p_{\epsilon, X}(y - \mu(x, \beta), x)$$

with the restriction $\mathbb{E}[\epsilon|X] = 0$.

Step 11 (cont.)

- Write

$$p_{\epsilon, X}(\epsilon, x) = \eta_1(\epsilon, x)\eta_2(x)$$

where $\eta_1(\epsilon, x) \equiv p_{\epsilon|X}(\epsilon|x)$ is non-negative and satisfies, for all x

$$\int \eta_1(\epsilon, x) d\epsilon = 1 \qquad \int \epsilon \eta_1(\epsilon, x) d\epsilon = 0$$

and $\eta_2(x) \equiv p_X(x)$ is non-negative and satisfies

$$\int \eta_2(\epsilon, x) dx = 1.$$

This defines the semiparametric model. Suppose the true functions are

$$\eta_{10}(\epsilon, x) \qquad \eta_{20}(x).$$

- Parametric submodel is

$$\mathcal{P}_{\beta,\gamma} \equiv \{p_Z(z, \beta, \gamma) = p_{\epsilon|X}(y - \mu(x, \beta)|x, \gamma_1)p_X(x, \gamma_2)\}$$

where γ_1 is $r_1 \times 1$ and γ_2 is $r_2 \times 1$, with $r = r_1 + r_2$.

Denote the true model

$$p_0(z) = p_{\epsilon|X}(y - \mu(x, \beta)|x, \gamma_{10})p_X(x, \gamma_{20}).$$

We have in the parametric submodel

$$\log p_Z(z, \beta, \gamma) = \log p_{\epsilon|X}(y - \mu(x, \beta)|x, \gamma_1) + \log p_X(x, \gamma_2).$$

So therefore

$$S_{\gamma_1}(z, \beta_0, \gamma_0) = \frac{\partial}{\partial \gamma_1} \{ \log p_{\epsilon|X}(y - \mu(x, \beta)|x, \gamma_1) \}_{\gamma_1 = \gamma_{10}}$$

$$S_{\gamma_2}(z, \beta_0, \gamma_0) = \frac{\partial}{\partial \gamma_2} \{ \log p_X(x, \gamma_2) \}_{\gamma_2 = \gamma_{20}}$$

Step 12 (cont.)

Suppressing the dependence of the score functions on the true values of the parameters, we have that a typical element in the parametric submodel nuisance tangent space is given by

$$BS_{\gamma}(\epsilon, X) = B_1 S_{\gamma_1}(\epsilon, X) + B_2 S_{\gamma_2}(X)$$

with B_1, B_2 $q \times r_1$ and $q \times r_2$ respectively. That is, in terms of the corresponding spaces

$$\Lambda_{\gamma} = \Lambda_{\gamma_1} \oplus \Lambda_{\gamma_2}$$

with Λ_{γ_1} and Λ_{γ_2} orthogonal as

$$\mathbb{E}[S_{\gamma_1}(\epsilon, X)\{S_{\gamma_2}(X)\}^{\top}] = 0_{r_1 \times r_2}$$

by properties of score functions and using iterated expectation.

Let

- Λ be the mean-square closure of all parametric submodel nuisance tangent spaces

$$\Lambda \equiv \{\text{mean-square closure of all } \Lambda_{\gamma_1} \oplus \Lambda_{\gamma_2}\};$$

- $\Lambda_{1s} = \{\text{mean-square closure of all } \Lambda_{\gamma_1}\};$
- $\Lambda_{2s} = \{\text{mean-square closure of all } \Lambda_{\gamma_2}\};$

Then

- Λ_{2s} is the space of all q -dimensional, mean-zero functions with finite variance (here the restriction is only that elements in the space are eligible score functions);
- Λ_{1s} is the space of all q -dimensional random functions $a(\epsilon, X)$ such that
 - (i) $\mathbb{E}[a(\epsilon, X)|X] = 0_{q \times 1}$;
 - (ii) $\mathbb{E}[a(\epsilon, X)\epsilon^\top | X] = 0_{q \times d}$

(as here the restriction is that we need a conditional mean zero given X , and

In the restricted mean model, the space orthogonal to the nuisance tangent space is

$$\{A(X)\epsilon : \text{where } A \text{ is } q \times d\}$$

and the projection of an arbitrary h onto this space is

$$h - \mathbb{E}[h\epsilon^\top | X] \left\{ \mathbb{E}[\epsilon\epsilon^\top | X] \right\}^{-1} \epsilon.$$

Step 12 (cont.)

The efficient score is then

$$\begin{aligned} S_{\text{eff}}(\epsilon, X) &= S_{\beta}(\epsilon, X) - \Pi(S_{\beta}(\epsilon, X) | \Lambda) \\ &\equiv \mathbb{E}[S_{\beta}(\epsilon, X) \epsilon^{\top} | X] \left\{ \mathbb{E}[\epsilon \epsilon^{\top} | X] \right\}^{-1} \epsilon = D(X)^{\top} \{V(X)\}^{-1} \epsilon \end{aligned}$$

say, where

$$D(X) = \left. \frac{\partial \mu(X, \beta)}{\partial \beta^{\top}} \right|_{\beta = \beta_0}$$

Hence we must (in principle) solve the estimating equation

$$\sum_{i=1}^n D(X_i)^{\top} \{V(X_i)\}^{-1} (Y_i - \mu(X_i, \beta)) = 0_{q \times 1}.$$

This optimal version depends on knowledge of the true model $p_0(\cdot)$.

In practice we cannot implement this, but inspired by this construction we can implement estimation based on

$$\sum_{i=1}^n A(X_i)(Y_i - \mu(X_i, \beta)) = 0_{q \times 1}.$$

where $A(X_i)$ is $q \times 1$.