WINPEC Working Paper Series No.E1920
January 2020

Differentiability of the Conditional Expectation

Hisatoshi Tanaka

Waseda INstitute of Political EConomy
Waseda University
Tokyo, Japan

Differentiability of the Conditional Expectation

Hisatoshi Tanaka

January 24, 2020

Abstract

The differentiability of a random function $b \mapsto E(Y|G_b(X))$ is presented in this study. The function is shown to be differentiable if $G_b(X)$ is continuously distributed. The result is applied to the semiparametric single-index model $E(Y|X) = F(G_\beta(X))$, and a quick way to compute the efficiency bound for estimating β has been proposed.

JEL Classification: C14

Keywords: Binary Response Model; Continuity of Conditional Expectations; Semiparametric Efficiency Bound; Semiparametric Estimation; Single Index Model

School of Political Science and Economics, Waseda University. Nishiwaseda 1-6-1, Shinjuku-ku, Tokyo 169-8050, Japan. E-mail: hstnk@waseda.jp

1 Introduction

This paper aims to find a sufficient condition for the differentiability of a random function,

$$b \mapsto E(Y|G_b(X)), \tag{1.1}$$

where (Y, X) is a random vector, $E(\cdot|\cdot)$ is a conditional expectation, and $G_b(\cdot)$ is a function indexed by $b \in \mathbb{R}^k$. The smoothness of (1.1) is a key assumption to derive the asymptotic variance and the efficiency bound of the semiparametric single index model (*e.g.* Cosslett (1987), Klein and Spady (1993), Sherman (1993), Chen and Lee (1998), Chen (2000), Song (2012) and Song (2014)). In most studies, the differentiability of (1.1) is often simply assumed, or proven under technical assumptions on the smoothness of the underlying probability densities.

The differentiability of (1.1) is not trivial. Crimaldi (2004) gives a quick example of a discontinuous conditional expectation: let X be an $\mathbb{R}$-valued random variable, $b \in \mathbb{R}$, and $f : \mathbb{R} \mapsto \mathbb{R}$ be a measurable function such that $Var(f(X)) > 0$. Then,

$$\gamma : b \mapsto E\left(f(X) \,|\, bX\right) \tag{1.2}$$

is not differentiable at $b = 0$ because $\gamma_b = E\left(f(X)|X\right) = f(X)$ for $b \neq 0$, while $\gamma_0 = Ef(X)$ at $b = 0$. Therefore, $\lim_{b \to 0} E(\gamma_b - \gamma_0)^2 \geq Var(f(X)) > 0$. The example is generalized to a higher dimensional case, which is presented as follows.

Proposition 1 Let $X = (X_1, X_2, \cdots, X_d)$ be a random vector, and let f be a measurable function of X such that $E\,Var(f(X)|X_2, \cdots, X_d) > 0$. Assume that $supp\,X_1$ is bounded, $supp\,(X_2, \cdots, X_d)$ is at most countable, and $\partial(supp\,(X_2, \cdots, X_d)) = \emptyset$. Then, $b \mapsto E\left(f(X) \,|\, X^\top b\right)$ is not continuous at $b = (0, b_2, \cdots, b_d) \in \mathbb{R}^d$.

Proof See Appendix A. ■

According to the proposition, $b \mapsto E(Y|G_b(X))$ is not differentiable at β if $supp\, G_b(X)$ changes in a discontinuous manner as $b \to \beta$. It is reasonable to guess that $E(Y|G_b(X))$ might become smooth at β if $G_\beta(X)$ is continuously distributed. The guess is proven correct by the paper.

The paper is organized as follows. Assumptions, definitions, and the main results of the paper are presented in section 2. An application of the results to semiparametric estimation is presented in section 3, where a quick way to compute the semiparametric efficiency bound of the single-index model is proposed. Section 4 concludes the paper.

2 Main Results

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space; $L_2(\mathbf{P})$ is the linear space of square integrable random variables on $(\Omega, \mathcal{F}, \mathbf{P})$ with the inner product $\langle W_1, W_2\rangle = \langle W_1, W_2\rangle_{\mathbf{P}} = \mathbf{P}(W_1 W_2)$ and the norm $\|W\| = \|W\|_{\mathbf{P}} = \langle W, W\rangle^{1/2}$; $L_2^0(\mathbf{P})$ is a set of $W \in L_2(\mathbf{P})$ such that $\mathbf{P}W = 0$; $L_2(Z)$ denotes a set of Z-measurable random variables; $L_2^0(Z) = \{W \in L_2(G_b(X)) : \mathbf{P}W = 0\}$.

Suppose that U is a nonempty open subset of $\mathbb{R}^k$ and $\mathcal{X} \subset \mathbb{R}^d$; $G_b : \mathcal{X} \mapsto \mathbb{R}$ is a function on $\mathcal{X}$ indexed with $b \in U$; $(Y, X) \in \mathbb{R} \times \mathcal{X}$ is a random vector such that

$$E(Y|X) = F(G_\beta(X)) \tag{2.1}$$

with $\beta \in U$, where $F : \mathbb{R} \mapsto \mathbb{R}$. Let $\mathcal{E}_b$ be the conditional-expectation operator defined by

$$\mathcal{E}_b Z = E(Z|G_b(X)) \tag{2.2}$$

for every $Z \in L_2(\mathbf{P})$; $\mathcal{E} = (\mathcal{E}_b)_{b\in U}$ generates a random function $\mathcal{E}Z : b \mapsto \mathcal{E}_b Z$. In the paper, the smoothness of the random function is defined as follows:

Definition 1 $\mathcal{W} : U \mapsto L_2(\mathbf{P})$, $b \mapsto \mathcal{W}_b$, is *continuous* if

$$\lim_{h\to 0} \|\mathcal{W}_{b+h} - \mathcal{W}_b\| = 0 \tag{2.3}$$

at every $b \in U$, and *differentiable* at β if there is $\partial_\beta \mathcal{W}_\beta \in (L_2(\mathbf{P}))^k$ such that

$$\left\|\mathcal{W}_{\beta+h} - \mathcal{W}_\beta - h^\top \partial_\beta \mathcal{W}_\beta\right\| = o(|h|) \tag{2.4}$$

for every $h \to 0$.

Throughout the paper, the following assumptions have been maintained:

(A1) At every $t \in \mathbb{R}$ and $b \in U$,

$$\lim_{h\to 0} \mathbf{P}\{G_{b+h}(X) = t\} = \mathbf{P}\{G_b(X) = t\}. \tag{2.5}$$

(A2) $G(X) : b \mapsto G_b(X)$ is almost surely continuous, that is,

$$\mathbf{P}\left\{\lim_{h\to 0} |G_{b+h}(X) - G_b(X)| = 0\right\} = 1 \tag{2.6}$$

for any $b \in U$, and differentiable at β with derivative $\partial_\beta G_\beta(X) \in (L_2(\mathbf{P}))^k$.

(A3) F is continuously differentiable on $\mathbb{R}$ with derivative $f = F'$.

Lemma 1 $\mathcal{E}Z : b \mapsto \mathcal{E}_b Z$ is continuous for any $Z \in L_2(\mathbf{P})$.

Proof Choose an arbitrary sequence $\{b_n\} \subset U$ such that $b_n \to b$ as $n \to \infty$. Let μ_n and μ be the laws of $W_n := G_{b_n}(X)$ and $W := G_b(X)$. The assumptions imply that $W_n \to W$ almost surely and that $\mu_n\{t\} \to \mu\{t\}$ for any $t \in \mathbb{R}$. By Theorem 1.6 of Crimaldi (2004), there exist $\{W'_n\}$ and $\{W'\}$ such that $\mathbf{P}\{W_n = W'_n\} = \mathbf{P}\{W = W'\} = 1$ for any $n \in \mathbb{N}$ and that $\|E(Z|W'_n) - E(Z|W')\| \to 0$ as $n \to \infty$. This implies existence of X' such that $\mathbf{P}\{X = X'\} = 1$ and that $\|E(Z|G_{b_n}(X)) - E(Z|G_b(X))\| = \|E(Z|G_{b_n}(X')) - E(Z|G_b(X'))\| \to 0$ as $n \to \infty$. ■

Theorem 1 $\mathcal{E}Y : b \mapsto \mathcal{E}_b Y$ is differentiable at β with derivative $(\partial_\beta \mathcal{E}_\beta)Y$, where

$$(\partial_\beta \mathcal{E}_\beta)Y = f(G_\beta(X))\Big[\partial_\beta G_\beta(X) - \mathcal{E}_\beta(\partial_\beta G_\beta(X))\Big]. \tag{2.7}$$

Remark 1 Let $\mathcal{E}_b^\perp = \mathrm{id} - \mathcal{E}_b$ be the orthogonal operator of $\mathcal{E}_b$, then the formula is shortly expressed as

$$(\partial_\beta \mathcal{E}_\beta)Y = \mathcal{E}_\beta^\perp \partial_\beta(\mathcal{E}_\beta Y). \tag{2.8}$$

Proof of Theorem 1 By the law of iterated expectations,

$$\mathcal{E}_b Y - \mathcal{E}_\beta Y = E(G_\beta(X)|G_b(X)) - G_\beta(X).$$

Let $\Delta_b := F(G_b(X)) - F(G_\beta(X)) - (b-\beta)^\top \partial_\beta F(G_\beta(X))$, where $\partial_\beta F(G_\beta(X)) = f(G_\beta(X))\partial_\beta G_\beta(X)$. Then,

$$\mathcal{E}_b Y - \mathcal{E}_\beta Y = \mathcal{E}_b^\perp \Big[(b-\beta)^\top \partial_\beta F(G_\beta(X)) + \Delta_b\Big]$$

and

$$\|\mathcal{E}_b Y - \mathcal{E}_\beta Y - (b-\beta)^\top \mathcal{E}_\beta^\perp \partial_\beta F(G_\beta(X))\| \le \|(\mathcal{E}_b - \mathcal{E}_\beta)\partial_\beta F(G_\beta(X))\| \cdot |b-\beta| + \|\mathcal{E}_b^\perp \Delta_b\|.$$

Lemma 1 implies $\|(\mathcal{E}_b - \mathcal{E}_\beta)\partial_\beta F(G_\beta(X))\| \to 0$ as $b \to \beta$. By the assumptions, $\|\mathcal{E}_b^\perp \Delta_b\| \le \|\Delta_b\| = o(|b-\beta|)$. Thus, $(\partial_\beta \mathcal{E}_\beta)Y = \mathcal{E}_\beta^\perp \partial_\beta F(G_\beta(X))$ is confirmed. ■

Corollary 1 Let $G_b(X) = X^\top b$ such that $Ef(X^\top\beta)^2 < \infty$ and $Var(X) < \infty$. If $X^\top b$ is continuously distributed on $\mathbb{R}$ for any b in the neighborhood of β,

$$\left.\frac{\partial}{\partial b} E(Y|X^\top b)\right|_{b=\beta} = f(X^\top\beta)\Big(X - E(X|X^\top\beta)\Big). \tag{2.9}$$

Proof The assumptions imply that $\lim_{b\to\beta} \mathbf{P}\{X^\top b = t\} = \mathbf{P}\{X^\top\beta = t\} = 0$ for every t and that $\|G_b(X) - G_\beta(X) - (b-\beta)^\top X\| = 0$. Hence, $\partial_\beta G_\beta(X) = X$ and $\mathcal{E}_\beta^\perp \partial_\beta(F(X^\top\beta)) = \mathcal{E}_\beta^\perp f(X^\top\beta)X = f(X^\top\beta)(X - \mathcal{E}_\beta X)$. ■

The continuous distribution of $X^\top b$ is often assumed for the identification of the linear-index model (*e.g.* Manski (1985), Manski (1988), Horowitz (1992)). Therefore, in most cases, the conditional expectation is presented as differentiable without assuming the technical assumptions on the density of (Y, X).

3 An Application to the Efficiency Bounds of the Single-Index Model

The right-hand side of the differentiation formula (2.7) displays its implications on the semiparametric efficiency bound of the single-index model. For example, consider a semiparametric regression model,

$$Y = F(G_\beta(X)) + u, \;\; E(u|X) = 0. \tag{3.1}$$

Assume $u \perp\!\!\!\perp X$ for the simplicity. Let $u \sim p$ and $X \sim q$. Assume that p satisfies *supp* $(p) = (-\infty, \infty)$, which is continuously differentiable on $\mathbb{R}$, and $\lim_{u\to\pm\infty} |p'(u)| = 0$.

Set $\varphi = \sqrt{p}$ and $\psi = \sqrt{q}$ according to the convention of the literature. Since $u = \mathcal{E}_\beta^\perp Y$, the log likelihood of parameter $\theta := (\mathcal{E}_\beta Y, \varphi, \psi)$ is given by

$$\ell_\theta = 2 \log \varphi(\mathcal{E}_\beta^\perp Y) + 2 \log \psi(X). \tag{3.2}$$

By Theorem 1, the derivative of ℓ_θ with respect to β is

$$\partial_\beta \ell_\theta = -\frac{2\varphi'(u)}{\varphi(u)} f(G_\beta(X)) \mathcal{E}_\beta^\perp(\partial_\beta G_\beta),$$

and

$$E[(\partial_\beta \ell_\theta)(\partial_\beta \ell_\theta)^\top]^{-1} = \frac{1}{4\|\varphi'\|_\lambda^2} E\left[f(G_\beta(X))^2 \, Var\Big(\partial_\beta G_\beta(X) \Big| G_\beta(X)\Big)\right]^{-1}, \tag{3.3}$$

where λ is the Lebesgue measure on $\mathbb{R}$ and $\|\varphi'\|_\lambda^2 = \int (\varphi')^2 d\lambda$. The right-hand side of (3.3) is identical to the efficiency bound for estimating β obtained by the standard method (*e.g.* Ichimura (1993), Severini and Tripathi (2013)).

The same holds for the standard binary response model,

$$Y = \{G_\beta(X) \geq u\}, \;\; u \perp\!\!\!\perp X, \;\; u \sim F, \;\; x \sim q. \tag{3.4}$$

The log-likelihood of $\theta = (\mathcal{E}_\beta Y, \psi)$ is proven by

$$\ell_\theta = [Y \log \mathcal{E}_\beta Y + (1 - Y) \log \mathcal{E}_\beta (1 - Y)] + 2 \log \psi(X), \tag{3.5}$$

and its derivative with respect to β is

$$\partial_\beta \ell_\theta = \frac{Y - F(G_\beta(X))}{F(G_\beta(X))(1 - F(G_\beta(X)))} f(G_\beta(X)) \mathcal{E}_\beta^\perp(\partial_\beta G_\beta).$$

Again,

$$E[(\partial_\beta \ell_\theta)(\partial_\beta \ell_\theta)^\top]^{-1} = E\left[\frac{f(G_\beta(X))^2 \, Var\,(\partial_\beta G_\beta(X) \,|\, G_\beta(X))}{F(G_\beta(X))(1 - F(G_\beta(X)))}\right]^{-1}$$

is identical to the known efficiency bound for estimating β in (3.4) obtained by the standard method (see *e.g.* Cosslett (1987)).

Under the standard parametrization, where the link function F is a nuisance parameter, treating F as known results in the incorrect evaluation of the efficiency bound due to the correlations between the score of β and the score of F. On the other hand, under the parametrization proposed in this paper, the efficiency bound for estimating β is directly obtained by $E[(\partial_\beta \ell_\theta)(\partial_\beta \ell_\theta)^\top]^{-1}$. This is because the score of β is already orthogonal to the score of $\mathcal{E}Y$. In Appendix B, the derivation of the efficiency bound for β of (3.1) is illustrated, and the reason why the parametrization by $\mathcal{E}_\beta Y$ works is explained.

4 Conclusions

This paper offers a set of sufficient conditions for the differentiability of the conditional expectation. Not only the smoothness of functional factors, but also the continuous distribution of the conditional variable is important for the differentiability. In semiparametric estimation, it is standard to assume the continuous distribution of at least one explanatory variable. Therefore, we can conclude that the conditional expectation is differentiable in most applications.

The differentiation formula of the conditional expectation has been applied to the efficiency bound for estimating the semiparametric single-index model. The efficiency bound of the model parametrized by the conditional expectation has been established to be equivalent to the bound of a finite dimensional model, where the nuisance parameter is assumed to be known. This equivalence might suggest that the parametrization by the conditional expectation is more natural for the single-index model than that by the link function.

References

Chen, S. (2000): "Efficient Estimation of Binary Choice Models under Symmetry," *Journal of Econometrics*, 96, 183–199.

Chen, S. and L.-f. Lee (1998): "Efficient Semiparametric Scoring Estimation of Sample Selection Models," *Econometric Theory*, 14, 423–462.

Cosslett, S. R. (1987): "Efficiency Bounds for Distribution-Free Estimators of the Binary Choice and the Censored Regression Models," *Econometrica*, 55, 559–585.

Crimaldi, I. (2004): "On the Behavior of the Conditional Expectations in Skorohod Representation Theorem," *Statistics & Probability Letters*, 67, 141–148.

Horowitz, J. L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505–531.

Ichimura, H. (1993): "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," *Journal of Econometrics*, 58, 71–120.

Klein, R. W. and R. H. Spady (1993): "An Efficient Semiparametric Estimator of Binary Response Models," *Econometrica*, 61, 387–421.

Manski, C. F. (1985): "Semiparametric Analysis of Discrete Response : Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 313–334.

——— (1988): "Identification of Binary Response Models," *Journal of the American Statistical Association*, 83, 729–738.

Severini, T. A. and G. Tripathi (2001): "A Simplified Approach to Computing Efficiency Bounds in Semiparametric Models," *Journal of Econometrics*, 102, 23–66.

——— (2013): *Semiparametric Efficiency Bounds for Microeconometric Models: A Survey*, vol. 6, Now Publishers, Inc.

Sherman, R. P. (1993): "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61, 123–137.

Song, K. (2012): "On the Smoothness of Conditional Expectation Functionals," *Statistics & Probability Letters*, 82, 1028–1034.

——— (2014): "Semiparametric Models with Single-Index Nuisance Parameters," *Journal of Econometrics*, 178, 471–483.

Appendix A Proof of Proposition 1

In the following, a proof for the case of $d = 2$ is specified for the simplicity of description. Let $\gamma_b := E(f(X)|X^\top b)$. Consider the case of $\theta > 0$. Assume that $supp\, X_1 = (0, 1)$ without any loss of generality. Let $supp\, X_2 = \{\xi_1 < \xi_2 < \xi_3 < \cdots\}$ and $\delta = \inf_j(\xi_{j+1} - \xi_j)$. Note that $\partial(supp\, X_2) = \emptyset$ implies $\delta > 0$. For any $t > 0$,

$$supp\, X^\top(t, \theta) \subseteq \bigcup_{j=1}^{\infty} (\theta\xi_j, \theta\xi_j + t). \tag{A.1}$$

If $0 < t < \theta\delta$, $(\theta\xi_i, \theta\xi_i + t) \cap (\theta\xi_j, \theta\xi_j + t) = \emptyset$ for $i \neq j$. Since $X^\top(t, \theta) \in (\theta\xi_j, \theta\xi_j + t)$ if and only if $X_2 = \xi_j$, $\sigma(X^\top(t, \theta)) = \sigma(X_1, X_2)$ for any $t \in (0, \theta\delta)$, which implies

$$\gamma_b = E(f(X)|X) = f(X) \tag{A.2}$$

for any $t \in (0, \theta\delta)$. On the other hand, $\gamma_{(0,\theta)} = E(f(X)|X_2)$ as $\theta > 0$. Therefore, $\lim_{t\downarrow 0} \|\gamma_{(t,\theta)} - \gamma_{(0,\theta)}\| = (E\, Var(f(X)|X_2))^{1/2} > 0$. The same argument holds for the case of $\theta < 0$.

Finally, in the case of $\theta = 0$, let $b_n = \left(\frac{1}{n^2}, \frac{1}{n}\right)$ for $n \in \mathbb{N}$. Then, for any $n > \frac{1}{\delta}$,

$$supp\, X^\top b_n \subseteq \bigcup_{j=1}^{\infty} \left(\frac{\xi_j}{n}, \frac{\xi_j}{n} + \frac{1}{n^2}\right) \tag{A.3}$$

with

$$\left(\frac{\xi_i}{n}, \frac{\xi_i}{n} + \frac{1}{n^2}\right) \cap \left(\frac{\xi_j}{n}, \frac{\xi_j}{n} + \frac{1}{n^2}\right) = \emptyset \tag{A.4}$$

for $i \neq j$. Therefore, $\lim_{n\to\infty} \|\gamma_{b_n} - \gamma_{(0,0)}\| = (\,Var f(X))^{1/2} > 0$.

∎

Appendix B The Efficiency Bound of the Single-Index Model Parametrized by the Conditional Expectation

In this appendix, the efficiency bound for estimating β of the semiparametric regression model (3.1) is derived, where the model is parametrized by $\theta = (\mathcal{E}_\beta Y, \varphi, \psi)$. The terms and concepts are according to Severini and Tripathi (2001). See also Severini and Tripathi (2013).

A key idea is to consider $\mathcal{E}_\beta Y$ as the value of the random function $\mathcal{E}Y : b \mapsto \mathcal{E}_b Y$ at β. Let $\mathcal{M}$ be a class of continuous random functions $\mathcal{W} : b \mapsto \mathcal{W}_b$ such that $\mathcal{W}_b \in L_2(G_b(X))$ and $\mathbf{P}\mathcal{W}_b = Y$ for every b. Let $\mathcal{M}_U := \bigcup_{b\in U}\{\mathcal{W}_b \,|\, \mathcal{W} \in \mathcal{M}\}$, then the parameter set is given by $\Theta = \mathcal{M}_U \times \Phi \times \Psi$, where λ is the Lebesgue measure on $\mathbb{R}$, μ is a reference measure on $\mathcal{X}$, $\Phi = L_2(\mathbb{R}, \lambda)$, and $\Psi = L_2(\mathcal{X}, \mu)$. The inner product $\langle\cdot,\cdot\rangle_\Theta$ is introduced on Θ by

$$\langle\theta_1, \theta_2\rangle_\Theta = \|\varphi'\|_\lambda^2 \langle M_1, M_2\rangle_{\mathbf{P}} + \langle\varphi_1, \varphi_2\rangle_\lambda + \langle\psi_1, \psi_2\rangle_\mu$$

for every $\theta_1 = (M_1, \varphi_1, \psi_1)$ and $\theta_2 = (M_2, \varphi_2, \psi_2)$, where $\langle\varphi_1, \varphi_2\rangle_\lambda = \int \varphi_1\varphi_2\, d\lambda$, and $\langle\psi_1, \psi_2\rangle_\mu = \int \psi_1\psi_2\, d\mu$. Let $\|\cdot\|_\Theta = \langle\cdot,\cdot\rangle_\Theta^{1/2}$ be the norm on Θ.

To find the tangent space of Θ at $\theta = (\mathcal{E}_\beta Y, \varphi, \psi)$, let $t \mapsto (b_t, \mathcal{W}_t)$ be a curve into $U \times \mathcal{M}$ such that

$$\left|\frac{b_t - \beta}{t} - h\right| = o(1) \quad \text{and} \quad \sup_{b\in U}\left\|\frac{\mathcal{W}_{t,b} - \mathcal{E}_b Y}{t} - e_b\right\| = o(1), \tag{B.1}$$

where $e : b \mapsto e_b \in L_2(G_b(X))$ is a continuous function and $h \in \mathbb{R}^k$. Let $t \mapsto \mathcal{W}_{t,b_t}$ be a curve into $\mathcal{M}_U$, and assume there exists a random variable $\dot{\mathcal{E}}_{\dot\beta} Y \in L_2(\mathbf{P})$ such that

$$\left\|\frac{\mathcal{W}_{t,b_t} - \mathcal{E}_\beta Y}{t} - \dot{\mathcal{E}}_{\dot\beta} Y\right\| = o(1). \tag{B.2}$$

The *tangent set* $T(\mathcal{E}_\beta Y, \mathcal{M}_U)$ of $\mathcal{M}_U$ at $\mathcal{E}_\beta Y$ is the set of $\dot{\mathcal{E}}_{\dot\beta} Y$ as in (B.2) for all curves $t \mapsto \mathcal{M}_{t,b_t}$, and the *tangent space* of $\mathcal{M}_U$ at $\mathcal{E}_\beta Y$ is given by $\overline{\operatorname{lin} T(\mathcal{E}_\beta Y, \mathcal{M}_U)}$. Then, the following lemma is expressed.

Lemma 2

$$T(\mathcal{E}_\beta Y, \mathcal{M}_U) = \overline{\operatorname{lin} T(\mathcal{E}_\beta Y, \mathcal{M}_U)} = L_2^0(G_\beta(X)) + \operatorname{span}\{(\partial_\beta \mathcal{E}_\beta)Y\}. \tag{B.3}$$

Proof Choose an arbitrary $\dot{\mathcal{E}}_{\dot\beta} Y \in T(\mathcal{E}_\beta Y, \mathcal{M}_U Y)$, then there exists a curve $t \mapsto (b_t, \mathcal{W}_t)$ such that (B.2) and that

$$\begin{aligned} \|\dot{\mathcal{E}}_{\dot\beta} Y - e_\beta - h^\top(\partial_\beta \mathcal{E}_\beta)Y\| \;&\le\; \left\|\frac{\mathcal{W}_{t,b_t} - \mathcal{E}_\beta Y}{t} - \dot{\mathcal{E}}_{\dot\beta} Y\right\| + \sup_{b\in U}\left\|\frac{\mathcal{W}_{t,b} - \mathcal{E}_b Y}{t} - e_b\right\| \\ &\quad + \|e_{b_t} - e_\beta\| + \left\|\left(\frac{\mathcal{E}_{b_t} - \mathcal{E}_\beta}{t} - h^\top(\partial_\beta\mathcal{E}_\beta)\right)Y\right\| \\ &= o(1). \end{aligned}$$

Since $\mathbf{P}\mathcal{W}_{t,\beta} = \mathbf{P}(\mathcal{E}_\beta Y) = \mathbf{P}Y$,

$$|\mathbf{P}e_\beta| = \left|\mathbf{P}\left(\frac{\mathcal{W}_{t,\beta} - \mathcal{E}_\beta Y}{t} - e_\beta\right)\right| = o(1),$$

which implies $\mathbf{P}e_\beta = 0$. Therefore, $\dot{\mathcal{E}}_{\dot\beta} Y = e_\beta + h^\top(\partial_\beta\mathcal{E}_\beta)Y \in L_2^0(G_\beta(X)) + \operatorname{span}\{(\partial_\beta\mathcal{E}_\beta)Y\}$ is shown.

Choose an arbitrary $e_\beta \in L_2^0(G_\beta(X))$ and $h \in \mathbb{R}^k$. Define a curve $t \mapsto \mathcal{W}_t$ in $\mathcal{M}$ by

$$\mathcal{W}_{t,b} := \mathcal{E}_b\left[Y + t\left(e_\beta + h^\top(\partial_\beta\mathcal{E}_\beta)Y\right)\right]$$

for every $b \in U$. By setting $b_t \equiv \beta$,

$$\left\|\frac{\mathcal{W}_{t,b_t} - \mathcal{E}_\beta Y}{t} - \left(e_\beta + h^\top(\partial_\beta\mathcal{E}_\beta)Y\right)\right\| = 0,$$

hence $e_\beta + h^\top(\partial_\beta\mathcal{E}_\beta)Y \in T(\mathcal{E}_\beta Y, \mathcal{M}_U)$ is shown. ∎

From the lemma, we have

$$\overline{\operatorname{lin} T(\theta, \Theta)} = (L_2^0(G_\beta(X)) + \operatorname{span}\{(\partial_\beta\mathcal{E}_\beta)Y\}) \times \varphi^\perp \times \psi^\perp, \tag{B.4}$$

where $\varphi^\perp = \{\dot\varphi \in L_2(\lambda) \,|\, \langle\dot\varphi, \varphi\rangle_\lambda = 0\}$ and $\psi^\perp = \{\dot\psi \in L_2(\mu) \,|\, \langle\dot\psi, \psi\rangle_\mu = 0\}$. Consider a curve $t \mapsto \theta_t = (\mathcal{W}_{t,b_t}, \varphi_t, \psi_t)$ into Θ passing through $\theta = (\mathcal{E}_\beta Y, \varphi, \psi)$ at $t = 0$. Let $\dot\theta = (e_\beta + h^\top(\partial_\beta\mathcal{E}_\beta)Y, \dot\varphi, \dot\psi) \in \overline{\operatorname{lin} T(\theta, \Theta)}$ be a tangent vector such that

$$\lim_{t\to 0}\left\|\frac{\theta_t - \theta}{t} - \dot\theta\right\|_\Theta = 0,$$

and let $\ell_t(X, Y) := \log\left[\varphi_t(Y - \mathcal{W}_{t,b_t})^2 \psi_t(X)^2\right]$ be the log-likelihood of the one-parameter submodel. The score for estimating $t = 0$ is

$$\dot\ell_0 = -\frac{2\varphi'(u)}{\varphi(u)}\left(e_\beta + h^\top(\partial_\beta\mathcal{E}_\beta)Y\right) + \frac{2\dot\varphi(u)}{\varphi(u)} + \frac{2\dot\psi(X)}{\psi(X)}.$$

The Fisher information for estimating $t = 0$ is given by

$$\mathbf{P}(\dot\ell_0)^2 = 4\|\varphi'\|_\lambda^2\,\|e_\beta + h^\top(\partial_\beta\mathcal{E}_\beta)Y\|_{\mathbf{P}}^2 + 4\|\dot\varphi\|_\lambda^2 + 4\|\dot\psi\|_\mu^2 = 4\|\dot\theta\|_\Theta^2.$$

The Fisher information metric $\langle\cdot,\cdot\rangle_F$ is thus $\langle\dot\theta_1,\dot\theta_2\rangle_F = 4\langle\dot\theta_1,\dot\theta_2\rangle_\Theta$. Since $L_2^0(G_\beta(X)) \perp \operatorname{span}\{(\partial_\beta\mathcal{E}_\beta)Y\}$ under $\langle\cdot,\cdot\rangle_{\mathbf{P}}$,

$$\langle\dot\theta_1,\dot\theta_2\rangle_F = 4\|\varphi'\|_\lambda^2 \left(\langle e_{1,\beta}, e_{2,\beta}\rangle_{\mathbf{P}} + h_1^\top V_\beta h_2\right) + 4\left\langle\dot\varphi_1,\dot\varphi_2\right\rangle_\lambda + 4\left\langle\dot\psi_1,\dot\psi_2\right\rangle_\mu,$$

where

$$V_\beta = E\left[\left((\partial_\beta\mathcal{E}_\beta)Y\right)\left((\partial_\beta\mathcal{E}_\beta)Y\right)^\top\right] = E\left[f(G_\beta(X))^2\, Var\left(\frac{\partial}{\partial\beta}G_\beta(X)\middle| G_\beta(X)\right)\right].$$

Define a functional $\rho_c : \Theta \mapsto \mathbb{R}$ by $\rho_c(\theta) = c^\top\beta$, where $c \in \mathbb{R}^k$ is arbitrary. The directional derivative of ρ_c is then

$$\nabla\rho_c(\dot\theta) := \left(\frac{d}{dt}\rho_c(\theta_t)\right)_{t=0} = c^\top h.$$

Let $\tilde\rho_c$ be the gradient of $\nabla\rho_c$ on $(\overline{\operatorname{lin} T(\theta,\Theta)}, \langle\cdot,\cdot\rangle_F)$ such that $\tilde\rho_c \in \overline{\operatorname{lin} T(\theta,\Theta)}$ and $\langle\tilde\rho_c,\dot\theta\rangle_F \equiv c^\top h$. The score equation is uniquely solved by $\tilde\rho_c = (\alpha^\top(\partial_\beta\mathcal{E}_\beta)Y, 0, 0)$ with $\alpha = (4\|\varphi'\|_\lambda^2)^{-1}V_\beta^{-1}c$. Hence, the efficiency bound for estimating $c^\top\beta$ is equal to

$$\|\tilde\rho_c\|_F^2 = \left(4\|\varphi'\|_\lambda^2\right)\alpha^\top V_\beta\alpha = c^\top\left[\left(4\|\varphi'\|_\lambda^2\right)^{-1}V_\beta^{-1}\right]c.$$

Since c is arbitrary, the efficiency bound for estimating β is $\left(4\|\varphi'\|_\lambda^2\right)^{-1}V_\beta^{-1}$, which is equal to (3.3).