Semiparametric Theory and Missing Data

Man Fai (Billy) IP, Di SU and Hon Kiu (James) TO

Department of Statistics, The Chinese University of Hong Kong

Chapter 4: Semiparametric Models

June 28, 2022

2021-22 Summer Reading Group

§1 GEE Estimators for the Restricted Moment Model

The Restricted Moment Model

In Chapter 3, we developed theoretical results for estimators of parameters in finite-dimensional parametric models where $Z_1, \ldots, Z_n$ are iid $\{p(z, \theta), \theta \in \Omega \subset \mathbb{R}^p\}$, p finite, and where θ can be partitioned as

$$\theta = \left(\beta^T, \eta^T\right)^T \quad \beta \in \mathbb{R}^q, \eta \in \mathbb{R}^r \quad p = q + r,$$

β being the parameter of interest and η the nuisance parameter.
In this chapter, we will consider semiparametric models that can be represented using the class of densities $p(z, \beta, \eta)$, where β, the parameter of interest, is finite-dimensional (q-dimensional); η, the nuisance parameter, is infinite-dimensional.

Definition: The restricted moment model considers the conditional expectation of a response variable Y given covariates X; that is, $\mathsf{E}(Y^{d\times 1}|X) = \mu^{d\times 1}(X, \beta)$ where d is the dimension of Y and β is q-dimensional. Equivalently,

$$Y_i = \mu(X_i, \beta) + \epsilon_i, \quad (1)$$

$$\mathsf{E}(\epsilon_i \mid X_i) = 0. \quad (2)$$

The true value of β is denoted by β_0.

- E.g. $d > 1$: multivariate and longitudinal response data as a function of covariates.
- E.g. $d = 1$: traditional regression models for a univariate response variable.

GEE Estimators

Definition: An example of a semiparametric estimator for the restricted moment model is the solution to the linear estimating equation

$$\sum_{i=1}^{n} A^{q\times d}\left(X_i, \hat{\beta}_n\right)\left\{Y_i^{d\times 1} - \mu^{d\times 1}\left(X_i, \hat{\beta}_n\right)\right\} = 0^{q\times 1}, \quad (3)$$

where $A\left(X_i, \beta\right)$ is an arbitrary $(q \times d)$ matrix of functions of the covariate X_i and the parameter β.

Such an estimator is an example of a solution to a generalized estimating equation, or GEE, as defined by Liang and Zeger (1986). It is also an example of an m-estimator as defined in Chapter 3.

Influence Functions of GEE Estimators

Influence Function

It holds that

$$n^{1/2}\left(\hat{\beta}_n - \beta_0\right) = n^{-1/2}\sum_{i=1}^{n}\left([E\{A(X,\beta_0)D(X,\beta_0)\}]^{-1}A(X_i,\beta_0)\right.$$
$$\left.\times\{Y_i - \mu(X_i,\beta_0)\}\right) + o_p(1), \quad (4)$$

where

$$D(X,\beta) = \frac{\partial \mu(X,\beta)}{\partial \beta^T} \quad (5)$$

is the gradient matrix $(d \times q)$. Consequently, the influence function for the i-th observation of $\hat{\beta}_n$ is

$$\{E(AD)\}^{-1}A(X_i,\beta_0)\{Y_i - \mu(X_i,\beta_0)\}, \quad (6)$$

where $A = A(X,\beta_0), D = D(X,\beta_0)$.

To find the influence function:

- By first order Taylor expansion, we have

$$0 = \sum_{i=1}^{n} A\left(X_i,\hat{\beta}_n\right)\left\{Y_i - \mu\left(X_i,\hat{\beta}_n\right)\right\} = \sum_{i=1}^{n} A(X_i,\beta_0)\{Y_i - \mu(X_i,\beta_0)\}$$
$$+\left\{\sum_{i=1}^{n}\mathcal{O}\left(Y_i,X_i,\beta_n^*\right) - \sum_{i=1}^{n} A\left(X_i,\beta_n^*\right)D\left(X_i,\beta_n^*\right)\right\}\left(\hat{\beta}_n - \beta_0\right), \quad (7)$$

and β_n^* denotes some intermediate value between $\hat{\beta}_n$ and β_0.

Influence Functions of GEE Estimators

To find the influence function (cont.):

- If we denote the rows of $A(X_i,\beta)$ by $\{A_1(X_i,\beta),\ldots,A_q(X_i,\beta)\}$, then $\mathscr{O}^{q\times q}(Y_i,X_i,\beta)$ is the $q\times q$ matrix defined by

$$\mathscr{O}^{q\times q}(Y_i,X_i,\beta)=\begin{pmatrix}\{Y_i-\mu(X_i,\beta)\}^T\,\frac{\partial A_1^T(X_i,\beta)}{\partial\beta^T}\\ \vdots\\ \{Y_i-\mu(X_i,\beta)\}^T\,\frac{\partial A_q^T(X_i,\beta)}{\partial\beta^T}\end{pmatrix}$$

  This matrix, although complicated, is made up of a linear combination of functions of X_i multiplied by elements of $\{Y_i-\mu(X_i,\beta)\}$.

- By (7), we obtain

$$n^{1/2}\left(\hat{\beta}_n-\beta_0\right)=\left\{-n^{-1}\sum_{i=1}^{n}\mathscr{O}\left(Y_i,X_i,\beta_n^*\right)+n^{-1}\sum_{i=1}^{n}A\left(X_i,\beta_n^*\right)D\left(X_i,\beta_n^*\right)\right\}^{-1}n^{-1/2}\sum_{i=1}^{n}A\left(X_i,\beta_0\right)\{Y_i-\mu(X,\beta_0)\}.$$

  The result follows from

$$n^{-1}\sum_{i=1}^{n}\mathscr{O}\left(Y_i,X_i,\beta_n^*\right)\xrightarrow{P}\mathsf{E}\{\mathscr{O}(Y,X,\beta_0)\}=0$$

  and

$$n^{-1}\sum_{i=1}^{n}A\left(X_i,\beta_n^*\right)D\left(X_i,\beta_n^*\right)\xrightarrow{P}\mathsf{E}\{A(X,\beta_0)D(X,\beta_0)\}.$$

Asymptotic Variance of GEE Estimators

As we demonstrated in Chapter 3, the asymptotic variance of an RAL estimator for β is the variance of its influence function. Therefore, the asymptotic variance of the GEE estimator $\hat{\beta}_n$ is the variance of (6).

Variance of Influence Functions of GEE Estimators

The asymptotic variance of $\hat{\beta}_n$ is

$$\{\mathsf{E}(AD)\}^{-1}\,\mathsf{E}\left\{AV(X)A^T\right\}\{\mathsf{E}(AD)\}^{-1^T}, \tag{8}$$

where $V(X_i) = \mathrm{Var}(Y_i \mid X_i)$ is the $d \times d$ conditional variance matrix of Y_i given X_i.

To find the asymptotic variance:

We first compute the variance of $A(X_i, \beta_0)\{Y_i - \mu(X_i, \beta_0)\}$, which equals

$$\begin{aligned}
\mathrm{Var}\left[A(X_i,\beta_0)\{Y_i-\mu(X_i,\beta_0)\}\right] =& \mathsf{E}\left(\mathrm{Var}\left[A(X_i,\beta_0)\{Y_i-\mu(X_i,\beta_0)\} \mid X_i\right]\right) \\
&+ \mathrm{Var}(\underbrace{\mathsf{E}\left[A(X_i,\beta_0)\{Y_i-\mu(X_i,\beta_0)\} \mid X_i\right]}_{=0}) \\
=&\mathsf{E}\left\{A(X_i,\beta_0)V(X_i)A^T(X_i,\beta_0)\right\},
\end{aligned} \tag{9}$$

Then

$$\begin{aligned}
&\mathrm{Var}\left(\{\mathsf{E}(AD)\}^{-1}A(X_i,\beta_0)\{Y_i-\mu(X_i,\beta_0)\}\right) \\
=&\{\mathsf{E}(AD)\}^{-1}\,\mathrm{Var}\left(A(X_i,\beta_0)\{Y_i-\mu(X_i,\beta_0)\}\right)\{\mathsf{E}(AD)\}^{-1^T}.
\end{aligned}$$

Estimation of the Asymptotic Variance

In order to use the results above for data analytic applications, such as constructing confidence intervals for β or for some components of β, we must also be able to derive consistent estimators for the asymptotic variance of $\hat{\beta}_n$ given by (7).

Estimator of the Asymptotic Variance (Assume β_0 known)

An estimator of the asymptotic variance is given by

$$\widetilde{\mathsf{Var}}(\sqrt{n}\hat{\beta}_n) = \hat{\mathsf{E}}_0(AD)^{-1}\hat{\mathsf{E}}_0(AVA^T)\hat{\mathsf{E}}_0(AD)^{-1T}, \tag{10}$$

where by the law of large numbers, a consistent estimator for $\mathsf{E}(AD)$ is given by

$$\hat{\mathsf{E}}_0(AD) = n^{-1}\sum_{i=1}^{n} A\left(X_i, \beta_0\right) D\left(X_i, \beta_0\right), \tag{11}$$

where the subscript "0" is used to emphasize that this statistic is computed with β_0 known. And by LLN, a consistent estimator of $\mathsf{E}\left\{A\left(X_i, \beta_0\right) V\left(X_i\right) A^T\left(X_i, \beta_0\right)\right\}$ is given by

$$\hat{\mathsf{E}}_0(AVA^T) = n^{-1}\sum_{i=1}^{n} A\left(X_i, \beta_0\right)\left\{Y_i - \mu\left(X_i, \beta_0\right)\right\}\left\{Y_i - \mu\left(X_i, \beta_0\right)\right\}^T A^T\left(X_i, \beta_0\right). \tag{12}$$

Estimation of the Asymptotic Variance

Estimator of the Asymptotic Variance (β_0 unknown)

Since $\hat{\beta}_n$ is a consistent estimator for β_0, a natural estimator for the asymptotic variance of $\hat{\beta}_n$ is given by

$$\widehat{\mathsf{Var}}(\sqrt{n}\hat{\beta}_n) = \{\hat{\mathsf{E}}(AD)\}^{-1}\hat{\mathsf{E}}(AVA^T)\{\hat{\mathsf{E}}(AD)\}^{-1T}, \tag{13}$$

where $\hat{\mathsf{E}}(AD)$ and $\hat{\mathsf{E}}(AVA^T)$ are computed as in equations (11) and (12), respectively, with $\hat{\beta}_n$ substituted for β_0.
This estimator is referred to as the sandwich estimator for the asymptotic variance. More details about this methodology can be found in Liang and Zeger (1986).

The results above did not depend on any specific parametric assumptions beyond the moment restriction and regularity conditions. Consequently, the estimator, given as the solution to equation (3), is a semiparametric estimator for the restricted moment model.

Example: Log-linear Model

The Log-linear Model

- Consider the problem where we want to model the relationship of a response variable Y ($d = 1$), which is positive, as a function of $q-1$ vector of covariates $X = (X_1, \ldots, X_{q-1})^T$. The log-linear model assumes that

$$\log\{\mathsf{E}(Y \mid X) = \alpha + \delta_1 X_1 + \cdots + \delta_{q-1} X_{q-1}\}.$$

  Here, the parameter of interest is given by $\beta^{q\times 1} = (\alpha, \delta_1, \ldots, \delta_{q-1})^T$.
- This is an example of a restricted moment model where

$$\mathsf{E}(Y|X) = \mu(X, \beta) = \exp(\alpha + \delta_1 X_1 + \cdots + \delta_{q-1} X_{q-1}). \quad (14)$$

- The log transformation guarantees that $\mathsf{E}(Y \mid X)$ is always positive. Consequently, this model puts no restrictions on the possible values that β can take.

GEE Estimator

With a sample of iid data $(Y_i, X_i), i = 1, \ldots, n$, if we take $A^{q\times 1}(X, \beta)$ in (3) to equal $(1, X_1, \ldots, X_{q-1})^T$, then the corresponding GEE estimator $\hat{\beta}_n = (\hat{\alpha}_n, \ldots, \hat{\delta}_{(q-1)n})^T$ is the solution to

$$\sum_{i=1}^{n} \left(1, X_i^T\right)^T \left\{Y_i - \exp\left(\alpha + \delta_1 X_{1i} + \ldots + \delta_{q-1} X_{(q-1)i}\right)\right\} = 0^{q\times 1}. \quad (15)$$

Example: Log-linear Model

Estimator of Asymptotic Variance

- The derivative matrix $D^{1\times q}(X,\beta)$, defined by (5), is equal to $\mu(X,\beta)\left(1, X^T\right)$
- Specifically,

$$\hat{E}(AD) = n^{-1}\sum_{i=1}^{n}\left(1, X_i^T\right)^T \mu\left(X_i, \hat{\beta}_n\right)\left(1, X_i^T\right),$$

$$\hat{E}\left(AVA^T\right) = n^{-1}\sum_{i=1}^{n}\left(1, X_i^T\right)^T \left\{Y_i - \mu\left(X_i, \hat{\beta}_n\right)\right\}^2 \left(1, X_i^T\right).$$

Remark 1. The asymptotic variance is the variance matrix of the limiting normal distribution to which $n^{1/2}\left(\hat{\beta}_n - \beta_0\right)$ converges. That is, the asymptotic variance is equal to the $(q\times q)$ matrix Σ, where $n^{1/2}\left(\hat{\beta}_n - \beta_0\right) \xrightarrow{\mathscr{D}} N(0,\Sigma)$. For practical applications, say, when we are constructing confidence intervals for δ_j, the regression coefficient for the j-th covariate $X_j, j = 1,\ldots,q-1$, we must be careful to use the appropriate scaling factor when computing the estimated standard error for $\hat{\delta}_{jn}$. That is, the 95% confidence interval for δ_j is given by

$$\hat{\delta}_{jn} \pm 1.96\,\mathrm{se}\left(\hat{\delta}_{jn}\right),$$

and $\mathrm{se}\left(\hat{\delta}_{jn}\right) = n^{-1}(\hat{\Sigma})_{(j+1)(j+1)}$, where $(\cdot)^{q\times q}_{(j+1)(j+1)}$ denotes the $(j+1)$-th diagonal element of the $q\times q$ matrix $(\cdot)^{q\times q}$.

§2 Parametric Submodels

Parametric Submodel

Recall in a semiparametric model, the data $Z_1, \ldots, Z_n$ are iid random vectors with a density that belongs to the class

$$\mathscr{P} = [p\{z, \beta, \eta(\cdot)\}, \text{where } \beta \text{ is } q\text{-dimensional and } \eta(\cdot) \text{ is infinite-dimensional}]$$

with respect to some dominating measure ν_Z. The infinite-dimensional nuisance parameter η is itself often a function and hence denoted as $\eta(\cdot)$. Denote the DGP by $p_0(z) \in P$, namely

$$p_0(z) = p\left\{z, \beta_0, \eta_0(\cdot)\right\}.$$

Motivation

Infinite-dimensional problems are tackled by first working with a finite-dimensional problem as an approximation and then taking limits to infinity.

Definition

A parametric submodel, which we will denote by $\mathscr{P}_{\beta,\gamma} = \{p(z, \beta, \gamma)\}$, is a class of densities characterized by the finite-dimensional parameter $\left(\beta^T, \gamma^T\right)^T$ such that

(i) $\mathscr{P}_{\beta,\gamma} \subset \mathscr{P}$ (i.e., every density in $\mathscr{P}_{\beta,\gamma}$ belongs to the semiparametric model $\mathscr{P}$) and

(ii) $p_0(z) \in \mathscr{P}_{\beta,\gamma}$ (i.e., the parametric submodel contains the truth). Another way of saying this is that there exists a density identified by the parameter (β_0, γ_0) within the parametric submodel such that

$$p_0(z) = p\left(z, \beta_0, \gamma_0\right).$$

In keeping with the notation of Chapter 3 , we will denote the dimension of γ by "r," although, in this case, the value r depends on the choice of parametric submodel.

Assumption and Remarks

- The parametric submodels that we will consider must satisfy certain smoothness conditions to allow the interchange of differentiation and integration of the density with respect to the parameters. This is necessary, for example, when we want to prove that the score vector has mean zero.
- Appropriate smoothness and regularity conditions on the likelihoods are given in Definition A.1 of the appendix in Newey (1990).
- A parametric model is a model whose probability densities are characterized through a finite number of parameters that the data analyst believes will suffice in identifying the DGP. For example, we may be willing to assume that our data follow the model

$$Y_i = \mu\left(X_i, \beta\right) + \varepsilon_i, \tag{16}$$

  where ε_i are iid $N\left(0, \sigma^2\right)$, independent of X_i. This model is contained within the semiparametric restricted moment model discussed previously.
- In contrast, a parametric submodel is a conceptual idea and we require a parametric submodel to contain the truth. But since we don't know what the truth is, we can only describe such submodels generically and hence such models are not useful for data analysis. The parametric model given by (16) is not a parametric submodel if, in truth, the data are not normally distributed.

Example: Proportional Hazards Model

The Proportional Hazards Model

Assume

$$\lambda(t \mid X) = \lambda(t) \exp\left(\beta^T X\right),$$

where $X = (X_1, \ldots, X_q)^T$ denotes a q-dimensional vector of covariates, $\lambda(t)$ is some arbitrary hazard function of time that is left unspecified and hence is infinite-dimensional, and β is the q-dimensional parameter of interest. Denote the truth by $\lambda_0(t); t \geq 0$ and β_0.

Its Parametric Submodel

Let $h_1(t), \ldots, h_r(t)$ be r different smooth functions of time that are specified by the data analyst. Consider the model $\mathscr{P}_{\beta,\gamma} = \{$class of densities with hazard function

$$\lambda(t \mid X) = \lambda_0(t) \exp\{\gamma_1 h_1(t) + \cdots + \gamma_r h_r(t)\} \exp\left(\beta^T X\right)\Big\},$$

where $\gamma = (\gamma_1, \ldots, \gamma_r)^T \in \mathbb{R}^r$ and $\beta \in \mathbb{R}^q$.

We note that:

- Parametric: In this model, the $(q + r)$ parameters $\left(\beta^T, \gamma^T\right)^T$ are left unspecified. Hence, this model is indeed a finite-dimensional model.
- Submodel: For any choice of β and γ, the resulting density follows a proportional hazards model and is therefore contained in the semiparametric model; i.e.,

  $$\mathscr{P}_{\beta,\gamma} \subset \mathscr{P}.$$

  The truth is obtained by setting $\beta = \beta_0$ and $\gamma = 0$.
- Not useful in practice: This parametric submodel is defined using $\lambda_0(t)$ which is not known.

§3 Influence Functions for Semiparametric RAL Estimators

IF of RAL estimators for parametric submodel

Recall that we study the geometry of the class of influence functions for RAL estimators for β for finite-dimensional parametric models.

Properties of influence functions of RAL estimators for β for a parametric submodel

Let $\varphi_{\beta,\gamma}(Z)$ be an influence function of RAL estimator for β for $\mathscr{P}_{\beta,\gamma}$.

1. $\varphi_{\beta,\gamma} \in \mathscr{W} \subseteq \mathscr{H}$, where $\mathscr{W} \perp \Lambda_\gamma = \{B^{q\times r} S_\gamma(Z,\beta_0,\gamma_0) \text{ for all } B^{q\times r}\}$.
2. **Efficient influence function** for the parametric submodel:
$$\varphi^{\text{eff}}_{\beta,\gamma}(Z) = \{\mathsf{E}(S^{\text{eff}}_{\beta,\gamma} S^{\text{eff}^{\mathrm{T}}}_{\beta,\gamma})\}^{-1} S^{\text{eff}}_{\beta,\gamma}(Z,\beta_0,\gamma_0),$$
where the **parametric submodel efficient score** is
$$S^{\text{eff}}_{\beta,\gamma}(Z,\beta_0,\gamma_0) = S_\beta(Z,\beta_0,\eta_0) - \Pi(S_\beta(Z,\beta_0,\eta_0) \mid \Lambda_\gamma).$$
3. The smallest asymptotic variance among such RAL estimators for β in the parametric submodel is
$$[\mathsf{E}\{S^{\text{eff}}_{\beta,\gamma}(Z) S^{\text{eff}^{\mathrm{T}}}_{\beta,\gamma}(Z)\}]^{-1}.$$

IF of RAL estimators for semiparametric model

Definition: An estimator for β is an RAL estimator for a **semiparametric model** if it is an RAL estimator for **every parametric submodels**.

This implies that

$$\{\text{Class of IF of RAL estimators for } \beta \text{ for } \mathscr{P}\} \subseteq \{\text{Class of IF of RAL estimators for } \beta \text{ for } \mathscr{P}_{\beta,\gamma}\}.$$

Heuristics: Suppose $\widehat{\beta}_n$ is a semiparametric estimator.

WTH: For all $p(z,\beta,\eta) \in \mathscr{P}$,

$$\sqrt{n}(\widehat{\beta}_n - \beta) \overset{D(\beta,\eta)}{\to} \mathrm{N}(0, \Sigma(\beta,\eta)). \tag{17}$$

For $\widehat{\beta}_n$ that satisfies (17), we must have

$$\sqrt{n}(\widehat{\beta}_n - \beta) \overset{D(\beta,\gamma)}{\to} \mathrm{N}(0, \Sigma(\beta,\gamma)), \tag{18}$$

for all $p(z,\beta,\gamma) \in \mathscr{P}_{\beta,\gamma} \subseteq \mathscr{P}$. However, (18) $\not\Rightarrow$ (17).

IF of RAL estimators for semiparametric model

Properties of influence functions of RAL estimators for β for a semiparametric model

Let $\varphi(Z)$ be an influence function of RAL estimator for β for $\mathscr{P}$.

1. $\varphi \perp \Lambda_\gamma$ for all $\mathscr{P}_{\beta,\gamma}$.
2. $\mathsf{Var}(\varphi) \geq [\mathsf{E}\{S^{\mathsf{eff}}_{\beta,\gamma}(Z) S^{\mathsf{eff}\,\mathrm{T}}_{\beta,\gamma}(Z)\}]^{-1}$ for all $\mathscr{P}_{\beta,\gamma}$. Hence,

$$\mathsf{Var}(\varphi) \geq \sup_{\mathscr{P}_{\beta,\gamma} \subseteq \mathscr{P}} \{\mathsf{E}(S^{\mathsf{eff}}_{\beta,\gamma} S^{\mathsf{eff}\,\mathrm{T}}_{\beta,\gamma})\}^{-1}.$$

Definition:

1. The **semiparametric efficiency bound** is defined as

$$\sup_{\mathscr{P}_{\beta,\gamma} \subseteq \mathscr{P}} \{\mathsf{E}(S^{\mathsf{eff}}_{\beta,\gamma} S^{\mathsf{eff}\,\mathrm{T}}_{\beta,\gamma})\}^{-1}. \tag{19}$$

2. A semiparametric RAL estimator $\widehat{\beta}_n$ with asymptotic variance achieving (19) for $p_0(z) = p(z, \beta_0, \eta_0)$ is said to be **locally efficient** at $p_0(\cdot)$.
3. A semiparametric RAL estimator $\widehat{\beta}_n$ is **globally semiparametric efficient** if it is semiparametric efficient regardless of $p_0(\cdot) \in \mathscr{P}$.

§4 Semiparametric Nuisance Tangent Space

Score function of semiparametric model

Definition: The **parametric submodel nuisance tangent spaces** is the set of elements $\{BS_\gamma(Z,\beta_0,\eta_0)\}$ and the mean-square closure of the above space, i.e. named as **semiparametric model nuisance tangent space** is defined as the $\Lambda \subset \mathscr{H}$, i.e.

$$\Lambda = \left[h(Z) \in \mathscr{H} s.t. \mathsf{E}\{h^{\mathrm{T}}(Z)h(Z)\} < \infty ; \exists B_j S_{\gamma_j}(Z) s.t. \|h(Z) - B_j S_{\gamma_j}(Z)\|^2 \xrightarrow{j\to\infty} 0 \right]$$

for a sequence of parametric submodels indexed by j.

Λ is larger and contains the union of all parametric submodel nuisance tangent spaces. Assume that Λ is a linear and closed space.

Definition: The **semiparametric efficient score** for β is defined as

$$S_{\text{eff}}(Z,\beta_0,\eta_0) = S_\beta(Z,\beta_0,\eta_0) - \prod\{S_\beta(Z,\beta_0,\eta_0)|\Lambda\}.$$

$\prod\{S_\beta(Z,\beta_0,\eta_0)|\Lambda\}$ exists and is unique since Λ is a closed linear subspace.

Theorem 4.1 Semiparametric efficiency bound

The **semiparametric efficiency bound** is equal to the inverse of the variance matrix of the semiparametric efficient score, i.e.

$$\left[\mathsf{E}\{S_{\text{eff}}(Z)S_{\text{eff}}^{\mathrm{T}}(Z)\}\right]^{-1}.$$

Proof of Theorem 4.1

Denote $\mathbf{V}$ as the semiparametric efficiency bound, i.e.

$$\sup_{\mathscr{P}_{\beta,\gamma}} \|S^{\text{eff}}_{\beta,\gamma}\|^{-2} = \mathbf{V}.$$

where $S^{\text{eff}}_{\beta,\gamma} = S_\beta(Z) - \prod(S_\beta(Z)|\Lambda_\gamma)$.
Since $\Lambda_\gamma \subset \Lambda$, this implies that $\|S_{\text{eff}}(Z)\| \le \|S^{\text{eff}}_{\beta,\gamma}(Z)\|$ for all parametric submodels $\mathscr{P}_{\beta,\gamma}$.
Hence,

$$\|S_{\text{eff}}\|^{-2} \ge \sup_{\text{all } \mathscr{P}_{\beta,\gamma}} \|S^{\text{eff}}_{\beta,\gamma}\|^{-2} = \mathbf{V}. \tag{20}$$

Since $\prod\{S_\beta(Z)|\Lambda\} \in \Lambda$, there exists a sequence of parametric submodels $\mathscr{P}_{\beta,\gamma_j}$ with nuisance score vectors $S_{\gamma_j}(Z)$ s.t.

$$\left\|\prod\{S_\beta(Z)|\Lambda\} - B_j S_{\gamma_j}(Z)\right\|^2 \xrightarrow{j\to\infty} 0.$$

By the definition of $\mathbf{V}$, for any $\mathscr{P}_{\beta,\gamma_j}$, therefore, $\mathbf{V}^{-1} \le \left\|S^{\text{eff}}_{\beta,\gamma_j}\right\|^2$ and obtain

$$\begin{aligned}
\mathbf{V}^{-1} &\le \left\|S_\beta(Z) - \prod(S_\beta(Z)|\Lambda_{\gamma_j})\right\|^2 \qquad \text{(Definition)}.\\
&\le \left\|S_\beta(Z) - B_j S_{\gamma_j}(Z)\right\|^2 \qquad \text{(Nuisance Tangent Space)}.\\
&= \left\|S_\beta(Z) - \prod(S_\beta(Z)|\Lambda)\right\|^2 + \left\|\prod(S_\beta(Z)|\Lambda) - B_j S_{\gamma_j}(Z)\right\|^2. \qquad (21)
\end{aligned}$$

From equation (21), by Pythagorean theorem, taking $j \to \infty$ imples

$$\left\|S_\beta(Z) - \prod(S_\beta(Z)|\Lambda)\right\|^2 = \|S_{\text{eff}}(Z)\|^2 \ge \mathbf{V}^{-1} \Leftrightarrow \|S_{\text{eff}}(Z)\|^{-2} \le \mathbf{V}.$$

Influence function of semiparametric RAL estimator (I/II)

Definition: The **efficient influence function** is defined as the influence function of a semiparametric RAL estimator, if

$$\mathsf{Var}(\varphi_{\text{eff}}) = \left\{\mathsf{E}(S_{\text{eff}} S_{\text{eff}}^{\mathrm{T}})\right\}^{-1}.$$

Theorem 4.2 Influence function for RAL estimator

Any semiparametric RAL estimator for β must have an **influence function** $\varphi(Z)$ that satisfies

1. $\mathsf{E}\{\varphi(Z)S_\beta^{\mathrm{T}}(Z,\beta_0,\eta_0)\} = \mathsf{E}\{\varphi(Z)S_{\text{eff}}^{\mathrm{T}}(Z,\beta_0,\eta_0)\} = I^{q\times q}$.
2. $\prod\{\varphi(Z)|\Lambda\} = 0$, i.e., $\varphi(Z)$ is orthogonal to the nuisance tangent space.

The efficient influence function is now defined as the unique element satisfying above conditions and it is equal to

$$\varphi_{\text{eff}}(Z,\beta_0,\eta_0) = \left\{\mathsf{E}(S_{\text{eff}} S_{\text{eff}}^{\mathrm{T}})\right\}^{-1} S_{\text{eff}}(Z,\beta_0,\eta_0).$$

The above Theorem results the semiparametric models can be parametrized through (β,η).

Proof of Theorem 4.2

- First prove condition (2). To show that $\varphi(Z)$ is orthogonal to $\Lambda \Leftrightarrow$ To show that $\langle \varphi, h \rangle = 0$ for all $h \in \Lambda$. By the definition of Λ, there exists a sequence $B_j S_{\gamma_j}(Z)$ s.t.

$$\left\| h(Z) - B_j S_{\gamma_j}(Z) \right\| \xrightarrow{j \to \infty} 0,$$

  for a sequence of parametric submodels indexed by j. Hence

$$\langle \varphi, h \rangle = \langle \varphi, B_j S_{\gamma_j} \rangle + \langle \varphi, h - B_j S_{\gamma_j} \rangle. \qquad \text{(Linearity of inner product)}.$$

  Since φ is orthogonal to Λ_{γ_j} (condition (ii) of Corollary 1 of Theorem 3.2), $\langle \varphi, B_j S_{\gamma_j} \rangle = 0$. Therefore,

$$|\langle \varphi, h \rangle| \leq \|\varphi\| \|h - B_j S_{\gamma_j}\| \qquad \text{(Cauch-Schwartz)}.$$

  As $j \to \infty$, $|\langle \varphi, h \rangle| \leq 0$.

- Then prove condition (1). Remark that $\varphi(Z)$ must satisfy $\mathsf{E}\{\varphi S_\beta^{\mathrm{T}}(Z, \beta_0, \eta_0)\} = I^{q \times q}$. Since

$$\mathsf{E}\{\varphi(Z) S_{\text{eff}}^{\mathrm{T}}(Z, \beta_0, \eta_0)\} = \mathsf{E}\{\varphi(Z) S_\beta^{\mathrm{T}}(Z, \beta_0, \eta_0)\} - \mathsf{E}\{\varphi(Z) \prod (S_\beta^{\mathrm{T}}(Z, \beta_0, \eta_0) | \Lambda)\},$$

  $\prod (S_\beta^{\mathrm{T}}(Z, \beta_0, \eta_0) | \Lambda) \in \Lambda$, and condition (2), then

$$\mathsf{E}\{\varphi(Z) \prod (S_\beta^{\mathrm{T}}(Z, \beta_0, \eta_0) | \Lambda)\} = 0.$$

Influence function of semiparametric RAL estimator (II/II)

Similarily, for semiparametric models, we could show the linear variety of influence functions of semiparametric RAL estimator and its corresponding efficient influence function as Chapter 3.

Theorem 4.3 Spaces and decomposition of influence function for RAL estimator

1. If a semiparametric RAL estimator for β exists, then the influence function of this estimator must belong to the space of influence functions, the **linear variety**

$$\left\{\varphi(Z) + \mathscr{T}^{\perp}\right\},$$

where $\varphi(Z)$ is the influence functions for any semiparametric RAL estimator for β and $\mathscr{T}$ is the semiparametric tangent space.

2. If an RAL estimator for β exists that achieves the semiparametric efficiency bound, then the influence function of this estimator must be the **unique and well-defined element**, i.e.

$$\varphi_{\text{eff}}(Z) = \varphi(Z) - \prod\{\varphi(Z)|\mathscr{T}^{\perp}\} = \prod\{\varphi(Z)|\mathscr{T}\}.$$

We are not clear whether there exist semiparametric estimators that will have influence functions corresponding to the elments of the Hillbert space satisfying Theorem 4.2 or 4.3. The construction of semiparametric estimators will be illustrated in the further section for semiparametric restricted moment model.

Tangent space for nonparametric models

Setup

- **Nonparametric model:** $Z_1, \ldots, Z_n \overset{\text{iid}}{\sim} p(z)$ with respect to a dominating measure ν_Z, where
$$p(z) \geq 0, \qquad \int p(z)\,\mathrm{d}\nu(z) = 1.$$
- **Parameter of interest:** $\beta \in \mathbb{R}^q$.
- **Hilbert space:**
$\mathscr{H} = \{$All q-dimensional mean-zero finite variance measurable functions$\}$, equipped with the covariance inner product.

Theorem 4.4 (Tangent space for nonparametric models)

$\mathscr{T} = \mathscr{H}$, i.e., the tangent space $\mathscr{T}$ (mean-square closure of all parametric submodel tangent spaces) is the entire Hilbert space $\mathscr{H}$.

Proof of Theorem 4.4 (1/3)

- **Step 1:** Consider parametric submodel $\mathscr{P}_\theta$.

  Let
  $$\mathscr{P}_\theta = \{p(z,\theta),\ \text{where } \theta \in \mathbb{R}^s\},$$
  be a parametric submodel. The **parametric submodel tangent space** is
  $$\Lambda_\theta = \{B^{q\times s} S_\theta(Z),\ \text{for all constant matrices } B^{q\times s}\},$$
  where
  $$S_\theta(z) = \frac{\partial \log p_0(z)}{\partial \theta}, \qquad p_0(z) = p(z,\theta_0).$$
  Suitable RCs imply
  $$\mathsf{E}_{\theta_0}\{S_\theta(Z)\} = 0^{s\times 1}.$$
  So, $\Lambda_\theta \subseteq \mathscr{H}$.

Man Fai (Billy) IP, Di SU and Hon Kiu (James) TO

Proof of Therorem 4.4 (2/3)

- **Step 2:** Show that any element of $\mathscr{H}$ can be written as an element of Λ_θ for some $\mathscr{P}_\theta$ or a limit of such elements.
  Choose an arbitrary bounded $h(Z) \in \mathscr{H}$.
  Consider the parametric submodel

$$p(z, \theta) = p_0(z)\{1 + \theta^{\mathrm{T}} h(z)\},$$

where $\theta \in \mathbb{R}^q$ is sufficiently small so that

$$\{1 + \theta^{\mathrm{T}} h(z)\} \geq 0 \qquad \text{for all } z. \tag{22}$$

Remarks:

1. Condition (22) is necessary to guarantee that $p(z, \theta)$ is a proper density.
2. Since $h(\cdot)$ is a bounded function, the set of θ satisfying (22) contains an open set in $\mathbb{R}^q$. This must be the case in order to ensure that the partial derivatives of $p(z, \theta)$ with respect to θ exist.

Note that every $p(z, \theta)$ in the parametric submodel satisfies

$$\begin{aligned}\int p(z, \theta)\, \mathrm{d}\nu(z) &= \int p_0(z)\{1 + \theta^{\mathrm{T}} h(z)\}\, \mathrm{d}\nu(z) \\ &= \underbrace{\int p_0(z)\, \mathrm{d}\nu(z)}_{=1} + \underbrace{\int \theta^{\mathrm{T}} h(z) p_0(z)\, \mathrm{d}\nu(z)}_{=0} = 1.\end{aligned}$$

Proof of Theorem 4.4 (3/3)

- **Step 2: (cont'd)**

  It guarantees that $p(z,\theta)$ is a proper density function for $\theta \in \mathscr{N}(\theta_0)$.
  The score vector for this parametric submodel is

$$S_\theta(z) = \left.\frac{\partial \log[p_0(z)\{1+\theta^{\mathrm{T}}h(z)\}]}{\partial\theta}\right|_{\theta=0} = h(z).$$

  Now, if we choose $B^{q\times q} = I^{q\times q}$, then

$$h(Z) = I^{q\times q}h(Z) \in \Lambda_\theta. \tag{23}$$

  (23) implies that $h(Z) \in \mathscr{T}$, i.e., the tangent space $\mathscr{T}$ contains all bounded mean-zero random vectors.
  The result follows by noting that any element of $\mathscr{H}$ can be approximated by a sequence of bounded h.

Basic setting

Setup

- **m-dimensional random vector:** $Z = (Z^{(1)}, \ldots, Z^{(m)})$.
- **Conditional density of $Z^{(j)}$ given $Z^{(1)}, \ldots, Z^{(j-1)}$ with respect to the dominating measure ν_j :**

$$p_{Z^{(j)} \mid Z^{(1)}, \ldots, Z^{(j-1)}}(z^{(j)} \mid z^{(1)}, \ldots, z^{(j-1)}). \tag{24}$$

- **Density of Z:**

$$p_Z(z) = p_{Z^{(1)}}(z^{(1)}) \prod_{j=2}^{m} p_{Z^{(j)} \mid Z^{(1)}, \ldots, Z^{(j-1)}}(z^{(j)} \mid z^{(1)}, \ldots, z^{(j-1)}).$$

Representation of nonparametric model

Representation by variationally independent nuisance parameters

If we put no restrictions on $p_Z(z)$ (nonparametric model), or equivalently, put no restrictions on (24), then the **jth conditional density** is any positive function $\eta_j(z^{(1)}, \ldots, z^{(j)})$ such that

$$\int \eta_j(z^{(1)}, \ldots, z^{(j)}) \, \mathrm{d}\nu_j(z^{(j)}) = 1,$$

for all $z^{(1)}, \ldots, z^{(j)}, j = 1, \ldots, m$.
Using this representation, the nonparametric model can be represented by the m variationally independent infinite-dimensional nuisance parameters $\eta_1(\cdot), \ldots, \eta_m(\cdot)$.

Definition: If the product of any combination of arbitrary $\eta_1(\cdot), \ldots, \eta_m(\cdot)$ can be used to construct a valid density for $Z^{(1)}, \ldots, Z^{(m)}$ in the nonparametric model, we say that $\eta_1(\cdot), \ldots, \eta_m(\cdot)$ are **variationally independent**.

Partition of parametric submodel tangent space

Recall that $\mathscr{T}$ is the mean-square closure of all parametric submodel tangent spaces.

Definitions:

1. A **parametric submodel** is given by the class of densities

$$p(z^{(1)}, \ldots, z^{(m)}, \gamma_1, \ldots, \gamma_m) = p(z^{(1)}, \gamma_1) \prod_{j=2}^{m} p(z^{(j)} \mid z^{(1)}, \ldots, z^{(j-1)}, \gamma_j), \tag{25}$$

   - $\gamma_j \in \mathbb{R}^{s_j}, j = 1, \ldots, m$: Variationally independent parameters
   - $p(z^{(j)} \mid z^{(1)}, \ldots, z^{(j-1)}, \gamma_{0,j})$: True conditional density of $Z^{(j)}$ given $Z^{(1)}, \ldots, Z^{(j-1)}$

2. The **parametric submodel tangent space** is defined as

$$\mathscr{T}_\gamma = \text{span}(\{S_{\gamma_j}(Z^{(1)}, \ldots, Z^{(m)}), j = 1, \ldots, m\}),$$

where

$$S_{\gamma_j}(z^{(1)}, \ldots, z^{(m)}) = \frac{\partial \log p(z^{(1)}, \ldots, z^{(m)}, \gamma_{0,1}, \ldots, \gamma_{0,m})}{\partial \gamma_j}.$$

Partition of parametric submodel tangent space

(25) $\Rightarrow$ the log-density is the sum of log-conditional densities with respect to variationally independent parameters γ_j, i.e.,

$$\log p(z^{(1)}, \dots, z^{(m)}, \gamma_1, \dots, \gamma_m) = \log p(z^{(1)}, \gamma_1) + \sum_{j=2}^{m} \log p(z^{(j)} \mid z^{(1)}, \dots, z^{(j-1)}, \gamma_j). \tag{26}$$

(26) implies that

$$\begin{aligned} S_{\gamma_j}(z^{(1)}) &= \frac{\partial \log p(z^{(1)}, \gamma_{0,1})}{\partial \gamma_1}, \\ S_{\gamma_j}(z^{(1)}, \dots, z^{(m)}) &= \frac{\partial \log p(z^{(j)} \mid z^{(1)}, \dots, z^{(j-1)}, \gamma_{0,j})}{\partial \gamma_j}, \qquad j = 2, \dots, m, \end{aligned}$$

is a function of $z^{(1)}, \dots, z^{(j)}$ only for $j = 2, \dots, m$.

Partition of parametric submodel tangent space

$$\mathscr{T}_\gamma = \bigoplus_{j=1}^{m} \mathscr{T}_{\gamma_j}, \tag{27}$$

where for $j = 1, \dots, m$,

$$\mathscr{T}_{\gamma_j} = \{B^{q \times s_j} S_{\gamma_j}(Z^{(1)}, \dots, Z^{(j)}) \text{ for all constant matrices } B^{q \times s_j}\}.$$

Partition of Hilbert space

Taking mean-square closure on (27), we have:

Partition of Hilbert space

$$\mathscr{T} = \bigoplus_{j=1}^{m} \mathscr{T}_j,$$

where for $j = 1, \ldots, m$, $\mathscr{T}_j$ is the mean-square closure of parametric submodel tangent spaces for $\eta_j(\cdot)$.

Partition of Hilbert space

Theorem 4.5 (Partition of Hilbert space)

$$\mathscr{T} = \mathscr{H} = \bigoplus_{j=1}^{m} \mathscr{T}_j, \tag{28}$$

where

$$\begin{aligned}
\mathscr{T}_1 &= \{\alpha_1^{q\times 1}(Z^{(1)}) \in \mathscr{H} : \mathsf{E}\{\alpha_1^{q\times 1}(Z^{(1)})\} = 0^{q\times 1}\}, \\
\mathscr{T}_j &= \{\alpha_j^{q\times 1}(Z^{(1)}, \ldots, Z^{(j)}) \in \mathscr{H} : \\
&\quad \mathsf{E}\{\alpha_j^{q\times 1}(Z^{(1)}, \ldots, Z^{(j)}) \mid Z^{(1)}, \ldots, Z^{(j-1)}\} = 0^{q\times 1}\}, \qquad j = 2, \ldots, m,
\end{aligned}$$

and $\mathscr{T}_j, j = 1, \ldots, m$ are mutually orthogonal spaces. In addition, any $h(Z^{(1)}, \ldots, Z^{(m)}) \in \mathscr{H}$ can be decomposed into orthogonal elements

$$h = \sum_{j=1}^{m} h_j,$$

where

$$\begin{aligned}
h_1(Z^{(1)}) &= \mathsf{E}\{h(\cdot) \mid Z^{(1)}\}, \\
h_j(Z^{(1)}, \ldots, Z^{(j)}) &= \mathsf{E}\{h(\cdot) \mid Z^{(1)}, \ldots, Z^{(j)}\} - \mathsf{E}\{h(\cdot) \mid Z^{(1)}, \ldots, Z^{(j-1)}\},
\end{aligned}$$

for $j = 2, \ldots, m$, and $h_j(\cdot)$ is the projection of h onto $\mathscr{T}_j$, i.e., $h_j(\cdot) = \Pi\{h(\cdot) \mid \mathscr{T}_j\}$.

Proof of Theorem 4.5 (1/4)

Step 1: Show that $\mathscr{T}_j$ is the mean-square closure of all parametric submodel tangent spaces for $\eta_j(\cdot)$ for $j = 1, \ldots, m$.

Recall that $S_{\gamma_j}(\cdot)$ is a function of $Z^{(1)}, \ldots, Z^{(j)}$ only, and under suitable RCs,

$$\mathsf{E}\{S_{\gamma_j}(Z^{(1)}, \ldots, Z^{(j)}) \mid Z^{(1)}, \ldots, Z^{(j-1)}\} = 0^{s_j \times 1}.$$

So, any q-dimensional element spanned by $S_{\gamma_j}(\cdot)$ must belong to $\mathscr{T}_j$.

Conversely, for any bounded element $\alpha_j(Z^{(1)}, \ldots, Z^{(j)}) \in \mathscr{T}_j$, consider the parametric submodel

$$p_j(z^{(j)} \mid z^{(1)}, \ldots, z^{(j-1)}, \theta_j) = p_{0j}(z^{(j)} \mid z^{(1)}, \ldots, z^{(j-1)})\{1+\theta_j^{\mathrm{T}} \alpha_j(z^{(1)}, \ldots, z^{(j)})\},$$

where

- $p_{0j}(z^{(j)} \mid z^{(1)}, \ldots, z^{(j-1)})$ denotes the true conditional density of $Z^{(j)}$ given $Z^{(1)}, \ldots, Z^{(j-1)}$, and
- θ_j is a q-dimensional parameter chosen sufficiently small to guarantee that $p_j(z^{(j)} \mid z^{(1)}, \ldots, z^{(j-1)}, \theta_j) \geq 0$.

Proof of Theorem 4.5 (2/4)

Step 1: (cont'd)

Notice that

$$\begin{aligned}\int p_{0j}(z^{(j)} \mid z^{(1)}, \ldots, z^{(j-1)})\, \mathrm{d}\nu_j(z^{(j)}) &= 1,\\ \int p_{0j}(z^{(j)} \mid z^{(1)}, \ldots, z^{(j-1)})\{\theta_j^{\mathrm{T}} \alpha_j(z^{(1)}, \ldots, z^{(j)})\}\, \mathrm{d}\nu_j(z^{(j)}) &= 0.\end{aligned}$$

The above implies that

$$\int p_{0j}(z^{(j)} \mid z^{(1)}, \ldots, z^{(j-1)})\{1 + \theta_j^{\mathrm{T}} \alpha_j(z^{(1)}, \ldots, z^{(j)})\}\, \mathrm{d}\nu_j(z^{(j)}) = 1,$$

verifying that the class of functions is a parametric submodel.
On the other hand, the score vector for the parametric submodel is

$$S_{\theta_j}(\cdot) = \left.\frac{\partial \log p_j(z^{(j)} \mid z^{(1)}, \ldots, z^{(j-1)}, \theta_j)}{\partial \theta_j}\right|_{\theta_j = 0} = \alpha_j(z^{(1)}, \ldots, z^{(j)}).$$

Up to this point, we have shown that

1. the tangent space for every parametric submodel of $\eta_j(\cdot)$ is contained in $\mathscr{T}_j$, and
2. every bounded element in $\mathscr{T}_j$ belongs to the tangent space for some parametric submodel of $\eta_j(\cdot)$.

The desired result follows by noting that every element of $\mathscr{T}_j$ is the limit of bounded elements of $\mathscr{T}_j$.

Proof of Theorem 4.5 (3/4)

Step 2: Show that $\mathscr{T}_j, j = 1, \ldots, m$ are mutually orthogonal spaces.
Take $j < j', h_j \in \mathscr{T}_j$ and $h_{j'} \in \mathscr{T}_{j'}$.
By the tower property, we have

$$\begin{aligned} \mathsf{E}(h_j^{\mathrm{T}} h_{j'}) &= \mathsf{E}\{\mathsf{E}(h_j^{\mathrm{T}} h_{j'} \mid Z^{(1)}, \ldots, Z^{(j'-1)})\} \\ &= \mathsf{E}\{h_j^{\mathrm{T}}\, \mathsf{E}(h_{j'} \mid Z^{(1)}, \ldots, Z^{(j'-1)})\} \\ &= 0. \end{aligned}$$

Proof of Theorem 4.5 (4/4)

Step 3: Verify the decomposition $h = h_1 + \cdots + h_m$.

It is clear that $h_j(\cdot) \in \mathscr{T}_j$, so by the projection theorem, it suffices to show that $h - h_j$ is orthogonal to every element of $\mathscr{T}_j$.

Consider any arbitrary $\ell_j \in \mathscr{T}_j$.

Since h_j and ℓ_j are functions of $Z^{(1)}, \ldots, Z^{(j)}$ by the tower property, we have

$$
\begin{aligned}
\mathsf{E}\{(h - h_j)^{\mathrm{T}} \ell_j\} &= \mathsf{E}[\mathsf{E}\{(h - h_j)^{\mathrm{T}} \ell_j \mid Z^{(1)}, \ldots, Z^{(j)}\}] \\
&= \mathsf{E}[\{\mathsf{E}(h \mid Z^{(1)}, \ldots, Z^{(j)}) - h_j\}^{\mathrm{T}} \ell_j] \\
&= \mathsf{E}[\{\mathsf{E}(h \mid Z^{(1)}, \ldots, Z^{(j-1)})\}^{\mathrm{T}} \ell_j] \qquad \text{(By definition of } h_j\text{)} \\
&= \mathsf{E}\left(\mathsf{E}[\{\mathsf{E}(h \mid Z^{(1)}, \ldots, Z^{(j-1)})\}^{\mathrm{T}} \ell_j \mid Z^{(1)}, \ldots, Z^{(j-1)}]\right) \\
&= \mathsf{E}[\{\mathsf{E}(h \mid Z^{(1)}, \ldots, Z^{(j-1)})\}^{\mathrm{T}} \mathsf{E}(\ell_j \mid Z^{(1)}, \ldots, Z^{(j-1)})] \\
&= 0 \qquad (\ell_j \in \mathscr{T}_j).
\end{aligned}
$$

Remark on Theorem 4.5

Remark: The linear space $\mathscr{T}_j$ can be equivalently defined as

$$\mathscr{T}_j = \{h_{\star j}^{q\times 1}(Z^{(1)},\ldots,Z^{(j)}) - \mathsf{E}\{h_{\star j}^{q\times 1}(Z^{(1)},\ldots,Z^{(j)}) \mid Z^{(1)},\ldots,Z^{(j-1)}\}\},$$

for all square-integrable functions $h_{\star j}^{q\times 1}$ of $Z^{(1)},\ldots,Z^{(j)}$.

Man Fai (Billy) IP, Di SU and Hon Kiu (James) TO