

ESTIMATION IN A SEMI-PARAMETRIC TWO-STAGE RENEWAL REGRESSION MODEL

Dorota M. Dabrowska

University of California, Los Angeles

Supplementary Material

This supplement has two sections. In Section S1, we give a derivation of the information bound for estimation of the Euclidean component of the model. In Section S2, we provide the proof of Lemma A.1 as well as some details of the Hoeffding expansion used in the proof of Propositions 3.1 and 3.2.

S1.Information bound

In the following the Euclidean parameter of interest is denoted by $\xi = (\beta, \theta)$, and we let $\eta = (A, A_{0c}, A_{1c}, A_{2c}, \mu)$ be the nuisance parameter. We let ξ_0 and $\eta_0 = (A^{(0)}, A_{0c}^{(0)}, A_{1c}^{(0)}, A_{2c}^{(0)}, \mu^{(0)})$ be the true parameter values. Without loss of generality, we assume that the cumulative hazards of the censoring distributions are absolutely continuous with respect to $(A_{ic}^{(0)}, i = 0, 1, 2)$ and denote the corresponding derivatives by $(\alpha_{ic}, i = 0, 1, 2)$. Likewise, we assume that the distribution μ of the covariates has density m with respect to the true distribution $\mu^{(0)}$.

The condition 2.1 (ii) implies that for $h \in \mathcal{H}_0^c$ we have $\mathbb{E}[Y_h(u)|Z = z] = S_0(u|z)G_0(u - |z)$, where $G_0(x|z)$ is the conditional survival function of C_1 given $Z = z$. Its cumulative hazard function is $A_{0c}^{(0)}$ and we have

$$Q_h(x_1, z) = \mathbb{E} N_h(x_1)1(Z \leq z) = \int_0^{x_1} \mathbb{E}[Y_h(u)|Z = z]A_h^{(0)}(du|z)\mu^{(0)}(dz)$$

for $h \in \mathcal{H}_0^c$. The condition 2 (iii) implies also that for $h \in \mathcal{H}_i^c, i = 1, 2$ we have $\mathbb{E}[Y_h(u)|Z = z, X_1 = x_1, J = i, \delta_1 = 1] = S_i(u|x_1, z)G_i(u - |x_1, z)$, where $G_i(u|x_1, z) = P(C_2 - X_1 > u|Z = z, J = i, X_1 = x, X_1 \leq C_1)$ is the survival function corresponding to $A_{ic}^{(0)}$. It follows that

$$Q_h(x_2, x_1, z) = \mathbb{E} N_h(x_2) N_{0i}(x_1) 1(Z \leq z) = \int_0^{x_2} \int_0^{x_1} \mathbb{E} Y_h(u) | X_1 = u_1, J = i, \delta_1 = 1, Z = z) A_h^{(0)}(du | u_1, z) Q_{0i}(d(u_1, z))$$

for $h \in \mathcal{H}_i^c$.

We shall need the following moment identities.

Lemma S1.1 Let $\varphi(W) = \{\varphi_h(X_1, Z_h), \varphi_{h'}(X_2, Z_{h'}) : h \in \mathcal{H}_0^c, h' \in \mathcal{H}_i^c, i = 1, 2\}$ be a vector of measurable functions. Define

$$I_h(\varphi) = \int \varphi_h(u, Z_h) M_h(du) \quad \text{if } h \in \mathcal{H}^c.$$

Then $I_h(\varphi) \in L_q(P)$, $q = 1, 2$ if and only if

$$\mathbb{E} \int |\varphi_h(u, Z_h)|^q N_h(du) = \mathbb{E} \int Y_h(u) |\varphi_h(u, Z_h)|^q A_h^{(0)}(du, Z_h) < \infty$$

for all $h \in \mathcal{H}^c$. In addition, $\mathbb{E} I_h(\varphi) = 0$, and

$$\begin{aligned} \mathbb{E} I_h(\varphi)^2 &= \mathbb{E} \int Y_h(u) \varphi_h(u, Z_h)^2 A_h^{(0)}(du | Z_h) \quad \text{if } h \in \mathcal{H}, \\ &= \mathbb{E} \int Y_h(u) \varphi_h(u, Z_h)^2 [1 - A_h^{(0)}(\Delta u | Z_h)] A_h^{(0)}(du | Z_h) \\ &\quad \text{if } h = (i, c), i = 0, 1, 2. \end{aligned}$$

For any $h, h' \in \mathcal{H}$ such that $h \neq h'$, we also have $\mathbb{E} I_h(\varphi) I_{h'}(\varphi) = 0$. \square

This lemma follows easily by noting that under conditions 2.1 the processes

$$\int_0^x \varphi_h(u, Z) M_h(du), x \geq 0, h \in \mathcal{H}_0^c$$

form orthogonal martingales with respect to the filtration $\mathcal{F}_{0x} = \sigma(N_h(u), Z, Y_h(u+) : u \leq x, h \in \mathcal{H}_0^c)$, and similarly, the processes

$$\int_0^x \varphi_h(u, Z, X_1) M_h(du), x \geq 0, h \in \mathcal{H}_i^c, i = 1, 2$$

form orthogonal martingales with respect to the filtration $\mathcal{F}_{ix} = \sigma(N_h(u), Y_h(u+), Z 1(\delta_1 = 1, J_1 = i), X_1 1(\delta_1 = 1, J_1 = i) : u \leq x, h \in \mathcal{H}_i^c)$. Using direct calculation, it is also easy to verify that the processes are orthogonal.

As in Nan, Edmond and Wellner (2004), in the following we assume that the censoring distributions are continuous. The condition 2.1 implies that each

subject contributes to the complete data log-likelihood the sum

$$L(\xi, \alpha) + L_c(\alpha_{ic}, i = 0, 1, 2) + L_Z(m),$$

where the three terms are given by

$$\begin{aligned} L(\xi, \alpha) &= \sum_{h \in \mathcal{H}} \int [\log q_h(Z_h, u, \theta) + r(Z, u, \beta)] N_h(du) + \\ &+ \sum_{h \in \mathcal{H}} \int \log \alpha(u) N_h(du) \\ &- \sum_{h \in \mathcal{H}} \int Y_h(u) q_h(Z_h, u, \theta) e^{r(Z, u, \beta)} \alpha(u) du, \\ L_c(\alpha_{ic}, i = 0, 1, 2) &= \sum_{i=0}^2 \int \log(\alpha_{ic}(du|Z_{ic})) N_{ic}(du) \\ &- \sum_{i=0}^2 \int Y_{ic}(u) \alpha_{ic}(u|Z_{ic}) A_{ic}^{(0)}(du|Z_{ic}), \\ L_Z(m) &= m(Z). \end{aligned}$$

The score function for the parameter of interest ξ and the nuisance parameter η are

$$\begin{aligned} \dot{\ell}_1^0[\xi](W) &= \sum_{h \in \mathcal{H}} \int \varphi_h(u, Z_h, \xi_0) M_h(du), \\ \dot{\ell}_2^0[a](W) &= \sum_{h \in \mathcal{H}} \int a(u) M_h(du), \quad a \in L_2\left(\sum_{h \in \mathcal{H}} (Q_h)\right), \\ \dot{\ell}_3^0[b_{0c}](W) &= \int b_{0c}(u, Z) M_{ic}(du), \quad b_{0c} \in L_2(Q_{0c}), \\ \dot{\ell}_4^0[b_{1c}](W) &= \int b_{1c}(u, Z_{1c}) M_{1c}(du), \quad b_{1c} \in L_2(Q_{1c}), \\ \dot{\ell}_5^0[b_{2c}](W) &= \int b_{2c}(u, Z_{2c}) M_{2c}(du), \quad b_{2c} \in L_2(Q_{2c}), \\ \dot{\ell}_6^0[c](W) &= c(Z), \quad c \in L_2^0(\mu^{(0)}). \end{aligned}$$

Here $\dot{\ell}_1^0[\xi]$ is the derivative of the log-likelihood with respect to the Euclidean parameter $\xi = (\theta, \beta)$, From Section 2, $\varphi_h(Z_h, u, \xi) = [\varphi_{1h}(Z_h, u, \xi), \varphi_{2h}(Z_h, u, \xi)]^T$ where

$$\varphi_{1h}(Z_h, u, \xi) = \dot{r}(Z, u, \beta),$$

$$\varphi_{2h}(Z_h, u, \xi) = \frac{\dot{q}_h}{q_h}(Z_h, u, \theta), \quad h \in \mathcal{H}.$$

For $h \in \mathcal{H}_0^c$, $Z_h = Z$, and we have

$$\begin{aligned} \varphi_{2h}(Z_h, u, \xi) &= \frac{\dot{q}_h}{q_h}(Z, u, \theta) && \text{for } h = (0, i), i = 1, 2, \\ &= -\frac{\dot{q}_1 + \dot{q}_2}{1 - q_1 - q_2}(Z, u, \theta) && \text{for } h = (0, 3). \end{aligned}$$

Further,

$$\begin{aligned} a(u) &= \frac{\partial}{\partial \varepsilon} \log \alpha_\varepsilon(u)|_{\varepsilon=0}, \\ b_{ic}(u, Z_{ic}) &= \frac{\partial}{\partial \eta_i} \log \alpha_{ic, \eta_i}(du|Z_{ic})|_{\eta_i=0}, i = 0, 1, 2, \\ c(Z) &= \frac{\partial}{\partial \kappa} \log m_\kappa(Z)|_{\kappa=0} \end{aligned}$$

for regular parametric submodels $\{\alpha_\varepsilon\}$, $\{\alpha_{ic, \eta_i}, i = 0, 1, 2\}$ and $\{m_\kappa\}$ passing through the parameters α , α_{ic} and m when $\varepsilon = \kappa = \eta_i = 0, i = 0, 1, 2$.

The tangent spaces are $\dot{\mathcal{P}}_i = [\dot{c}_i^0]$, $i = 1, \dots, 6$ and $[\alpha]$ is the closed linear span generated by the score α . Using Lemma S1.1, it is easy to see that the spaces $\dot{\mathcal{P}}_i, i = 2, \dots, 6$ are mutually orthogonal, and $\dot{\mathcal{P}}_1$ is orthogonal to $\dot{\mathcal{P}}_i, i = 3, \dots, 6$.

The following proposition generalizes Propositions 3.1 and 3.2 in Nan, Edmond and Wellner (2004).

Proposition S1.2

(i) Define

$$\dot{\mathcal{P}}^* = \left\{ \sum_{h \in \mathcal{H}^c} \int g_h(u, Z_h) M_h(du) g_0(Z) : g_h \in L_2(Q_h), h \in \mathcal{H}^c, g_0 \in L_2(m) \right\}.$$

Then $\dot{\mathcal{P}}^* = L_2(P)$.

(ii) Let $c(W) = \{c_h(X_1, Z_h), c_{h'}(X_2, Z_{h'}) : h \in \mathcal{H}_0, h' \in \mathcal{H}_i, i = 1, 2\}$ be a vector of functions such that $c_h \in L_2(Q_h)$ for each $h \in \mathcal{H}$. Define

$$B(c) = \sum_{h \in \mathcal{H}} \int [c_h - e_c](u) M_h(du)$$

by setting $e_c(u) = s_c^{(1)}(u)/s_0(u)$, where

$$s_c^{(1)}(u) = \sum_{h \in \mathcal{H}} \mathbb{E} Y_h(u) c_h(u, Z_h) q_h(u, Z_h, \theta_0) e^{r(Z, u, \beta_0)},$$

$$s_0(u) = \sum_h \mathbb{E} Y_h(u) q_h(u, Z_h, \theta_0) e^{r(Z, u, \beta_0)}.$$

Then $B(c) \perp \dot{P}_\eta \in L_2^0(P)$ and $\dot{P}_\eta^\perp = (\sum_{j=1}^6 \dot{P}_j)^\perp = \{B(c) : c_h \in L_2(Q_h), h \in \mathcal{H}\}$.

When specialized to the renewal model considered in the paper, this Proposition implies that the efficient score function of estimation of the parameter $\xi = (\theta, \beta)$ corresponds to the choice of $c(W) = \{\varphi_h(Z_h, u, \xi) : h \in \mathcal{H}\}$, which is quite similar to the standard Cox regression. Note, however, that the model assumes that the functions $r(Z, u, \beta)$ and $q_h(Z_h, u, \theta), h \in \mathcal{H}$ have support not depending on the unknown parameters. In the absence of covariates, this assumption fails for instance in the parametric Marshall- Olkin model (1967), where \sqrt{n} rate of convergence of the asymptotically efficient estimators applies only to the interior of the parameter set.

Proof. Recall that $Z_h = Z$, for $h \in \mathcal{H}_0^c$ and $Z_h = (X_1, Z)$ for $h \in \mathcal{H}_i^c, i = 1, 2$. Any function $f(W) \in L_2(P)$ can be represented as a sum

$$\begin{aligned} f(W) &= \delta_1 \delta_2 \sum_{j=1}^2 1(J_1 = j) f_{j3}(X_2, X_1, Z) \\ &+ \delta_1 (1 - \delta_2) \sum_{j=1}^2 1(J_1 = j) f_{jc}(X_2, X_1, Z) \\ &+ \delta_1 1(J_1 = 3) f_{03}(X_1, Z) + (1 - \delta_1) f_{0c}(X_1, Z). \end{aligned}$$

For $j = 1, 2$ define

$$f_{0j}(X_1, Z) = \mathbb{E} [\delta_2 f_{j3}(X_2, X_1, Z) + (1 - \delta_2) f_{jc}(X_2, X_1, Z) | X_1, Z, \delta_1 = 1, J = j].$$

Then

$$\begin{aligned} f(W) &= I_1(W) + I_2(W), \\ I_1(W) &= \delta_1 \sum_{j=1}^3 1(J = j) f_{0j}(X_1, Z) + (1 - \delta_1) f_{0c}(X_1, Z), \\ I_2(W) &= \delta_1 \delta_2 \sum_{j=1}^2 1(J = j) f_{j3}(X_1, X_2, Z) + (1 - \delta_2) \delta_1 \sum_{j=1}^2 1(J = j) f_{jc}(X_1, X_2, Z) \\ &- \delta_1 \sum_{j=1}^2 1(J = j) f_{0j}(X_1, Z). \end{aligned}$$

Set

$$\begin{aligned} R_1[f](x_1, z) &= \sum_{j=1}^3 \mathbb{E}[\delta_1 f_{0j}(X_1, Z) | X_1 > x_1, Z = z] \\ &+ \mathbb{E}[(1 - \delta_1) f_{0c}(X_1, Z) | X_1 > x_1, Z = z] \end{aligned}$$

and

$$\begin{aligned} R_{2j}[f](x_2, x_1, z) &= \mathbb{E}[\delta_2 f_{j3}(X_2, X_1, Z) | X_2 > x_2, X_1 = x_1, Z = z, J_1 = j, \delta_1 = 1] \\ &+ \mathbb{E}[(1 - \delta_2) f_{j1}(X_2, X_1, Z) | X_2 > x_2, X_1 = x_1, Z = z, J_1 = j, \delta_1 = 1] \end{aligned}$$

for $j = 1, 2$. Finally, let

$$\begin{aligned} g_0(Z) &= \mathbb{E} f(W) | Z, \\ g_h(X_1, Z) &= f_{0j}(X_1, Z) - R_1[f](X_1, Z), \quad h \in \mathcal{H}_0^c, \\ g_h(X_2, X_1, Z) &= f_h(X_2, X_1, Z) - R_{2j}[f](X_2, X_1, Z), \quad h \in \mathcal{H}_j^c, j = 1, 2. \end{aligned}$$

We have

$$\begin{aligned} R_1[f](x_1, Z) - g_0(Z) &= \frac{\sum_{h \in \mathcal{H}_0^c} \int_{x_1}^{\infty} f_h(u, Z) Q_h(du, Z)}{1 - \sum_{h \in \mathcal{H}_0^c} Q_h(du, Z)} - g_0(Z) \\ &= - \sum_{h \in \mathcal{H}_0^c} \int_0^{x_1} [f_h(u, Z) - R_1[f](u, Z)] A_h^{(0)}(du | Z). \end{aligned}$$

Therefore

$$\begin{aligned} I_1(W) &= \sum_{j=1}^3 \delta_1 \mathbf{1}(J = j) f_{0j}(X_1, Z) + (1 - \delta_1) f_{0c}(X_1, Z) - R_1[f](X_1, Z) \\ &+ \left(R_1[f](X_1, Z) - g_0(Z) \right) + g_0(Z) \\ &= \sum_{h \in \mathcal{H}_0^c} \int g_h(u, Z) M_h(du) + g_0(Z) = \sum_{h \in \mathcal{H}_0^c} \int g_h(u, Z_h) M_h(du) + g_0(Z) \end{aligned}$$

Similarly, for $j = 1, 2$,

$$\begin{aligned} R_{2j}[f](x_2, X_1, Z) - f_{0j}(X_1, Z) &= \\ &- \sum_{h \in \mathcal{H}_j^c} \int_0^{x_2} [f_h(u, X_1, Z) - R_{2j}[f](u, X_1, Z)] A_h^{(0)}(du | X_1, Z) \end{aligned}$$

and

$$\begin{aligned}
& \delta_1 \delta_2 1(J = j) f_{j1}(X_2, X_1, Z) + (1 - \delta_2) \delta_1 f_{jc}(X_1, X_2, Z) = \\
& = \delta_1 \delta_2 1(J = j) f_{j1}(X_2, X_1, Z) + (1 - \delta_2) \delta_1 f_{jc}(X_1, X_2, Z) - R_{2j}[f](x_2, X_1, Z) \\
& + \left(R_{2j}[f](x_2, X_1, Z) - f_{0j}(X_1, Z) \right) + f_{0j}(X_1, Z) \\
& = \sum_{h \in \mathcal{H}_j^c} \int g_h(u, X_1, Z) M_h(du) + f_{0j}(X_1, Z) \delta_1 1(J = j).
\end{aligned}$$

Summing over j we obtain

$$I_2(W) = \sum_{j=1}^2 \sum_{h \in \mathcal{H}_j^c} \int g_h(u, X_1, Z) M_h(du) = \sum_{j=1}^2 \sum_{h \in \mathcal{H}_j^c} \int g_h(u, Z_h) M_h(du).$$

It follows that $f(W) \in \dot{\mathcal{P}}^*$.

To show part (ii) of the proposition, let

$$b(W) = \sum_{h \in \mathcal{H}} \int c_h dM_h.$$

Then $\Pi(b|\dot{\mathcal{P}}_2) = \sum_h \int a^*(u) dM_h(u)$ for a square integrable function a^* such that

$$\mathbb{E} \left(\sum_{h \in \mathcal{H}} \int a dM_h \right) \left(\sum_{h \in \mathcal{H}} \int [c_h - a^*] dM_h \right) = 0$$

for any $a \in L_2(\sum_{h \in \mathcal{H}} Q_h)$. Orthogonality implies that the left side is equal to

$$\int (a s_c^{(1)} - a^* s^0) A(du)$$

so that $a^* = s_c^{(1)}/s^0$. The proof can be completed as in Nan et al. (2004) and using that $B(c)$ is orthogonal to $\sum_{j=3}^6 \dot{\mathcal{P}}_j$. \square

S2. Proof of Lemma A.1

Before showing this lemma, we recall that a class of functions \mathcal{G} defined on some measure space (Ω, \mathcal{A}) is Euclidean for envelope G if $|g| \leq G$ for all $g \in \mathcal{G}$, and there exist constants A and V such that $N(\varepsilon \|G\|_{L_2(P)}, \mathcal{G}, \|\cdot\|_{L_2(P)}) \leq (A/\varepsilon)^V$ for all $\varepsilon \in (0, 1)$ and all probability measures P such that $\|G\|_{L_2(P)} < \infty$. Here $\|\cdot\|_{L_2(P)}$ is the $L_2(P)$ norm and $N(\eta, \mathcal{G}, \|\cdot\|_{L_2(P)})$ is the minimal number of $L_2(P)$ -balls of radius η covering the class \mathcal{G} . In the case of classes \mathcal{G}_n changing with n , the Euclidean constants A and V are taken to be independent of n .

It is easy to verify that the condition $\mathbb{E} G(\mathbf{W}_m)^p < \infty$ implies

$$\begin{aligned} \mathbb{E} G^p(\mathbf{W}_m) 1(G(\mathbf{W}_m) \geq n^{m/p}) &\rightarrow 0, \\ n^{-m(2-p)/p} \mathbb{E} G^2(\mathbf{W}_m) 1(G(\mathbf{W}_m) < n^{m/p}) &\rightarrow 0. \end{aligned}$$

Let

$$\begin{aligned} g_{1n}(\mathbf{W}_m) &= g(\mathbf{W}_m) 1(G(\mathbf{W}_m) < n^{m/p}), \\ g_{2n}(\mathbf{W}_m) &= g(\mathbf{W}_m) 1(G(\mathbf{W}_m) \geq n^{m/p}). \end{aligned}$$

Since g is a canonical kernel, we have

$$\mathbb{U}_{n,m}(g) = \mathbb{U}_{n,m}(\pi_m[g_{1n}]) + \mathbb{U}_{n,m}(\pi_m[g_{2n}]).$$

Define

$$\begin{aligned} G_{1n}(\mathbf{W}_m) &= \sum_A \mathbb{E}_A G(\mathbf{W}_m) 1(G(\mathbf{W}_m) < n^{m/p}), \\ G_{2n}(\mathbf{W}_m) &= \sum_A \mathbb{E}_A G(\mathbf{W}_m) 1(G(\mathbf{W}_m) \geq n^{m/p}). \end{aligned}$$

Then $G_{1n}(\mathbf{W}_m)$ and $G_{2n}(\mathbf{W}_m)$ form envelopes for the classes of truncated canonical kernels $\mathcal{G}_{1n} = \{\pi_m[g_{1n}] : g \in \mathcal{G}\}$ and $\mathcal{G}_{2n} = \{\pi_m[g_{2n}] : g \in \mathcal{G}\}$, respectively.

We have

$$\begin{aligned} \mathbb{E} n^{m(p-1)/p} \sup_{g_{2n} \in \mathcal{G}_{2n}} |\mathbb{U}_{n,m}(\pi_m[g_{2n}])| &\leq n^{m(p-1)/p} \mathbb{E} G_{2n}(\mathbf{W}_m) \leq \\ C n^{m(p-1)/p} \mathbb{E} G(\mathbf{W}_m) 1(G(\mathbf{W}_m) \geq n^{m/p}) &\leq C \mathbb{E} G^p(\mathbf{W}_m) 1(G(\mathbf{W}_m) \geq n^{m/p}) \end{aligned}$$

and the right-hand side tends to 0. We also have

$$\mathbb{E} n^{m(p-1)/p} \sup_{g_{1n} \in \mathcal{G}_{1n}} |\mathbb{U}_{n,m}(\pi_m[g_{1n}])| = \mathbb{E} n^{m/2} \sup_{g_{1n} \in \mathcal{G}_{1n}} |\mathbb{U}_{n,m}(\pi_m[g_{1n}/n^{m(2-p)/2p}])|.$$

The class \mathcal{G}_{1n} is Euclidean for the envelope $G_{1n}(\mathbf{W}_m)$. By the randomization Theorem 3.5.3 and Corollary 5.1.8 of de la Peña and Giné (1999) or using similar developments as on page 254-255 of their text, the right-hand side of the above display is of order

$$\begin{aligned} &K \mathbb{E} \sqrt{\mathbb{U}_{n,m}(G_{1n}^2(\mathbf{W}_m)/n^{m(2-p)/p})} \leq K [\mathbb{E} [G_{1n}^2(\mathbf{W}_m)/n^{m(2-p)/p}]^{1/2} \\ &= KC [n^{-m(2-p)/p} \mathbb{E} [G^2(\mathbf{W}_m) 1(G(\mathbf{W}_m))]^{1/2} \rightarrow 0. \end{aligned}$$

Here K is a constant depending only on m and the VC-characteristics of the class \mathcal{G}_{1n} , but not n , while C is bounded by 2^m . This completes the proof. \square

Finally, we give some more details of the Hoeffding decomposition used in the proofs of Lemmas A.2 and A.3. In calculations given below, we use $A = A_0$, the true baseline cumulative hazard function.

In Lemma A.2, we consider first the statistic $\mathbb{U}_{n,2}(g^{(2)})$. For $i \neq j$, we have $\mathbb{E} g^{(2)}(W_i, W_j) = 0$,

$$\begin{aligned} \mathbb{E}_{\{1\}} g^{(2)}(W_i, W_j) &= \mathbb{E} [g^{(2)}(W_i, W_j) | W_i] = \sum_h \int_0^\tau \left[\frac{\varphi_{ih} s^{(0)} - s^{(1)}}{s^{(0)}} \right] (u, \xi_0) N_{hi}(du), \\ \mathbb{E}_{\{2\}} g^{(2)}(W_i, W_j) &= \mathbb{E} [g^{(2)}(W_i, W_j) | W_j] = \\ &= \sum_h \int_0^\tau \left[\frac{(s^{(1)} S_j^{(0)}) - S_j^{(1)} s^{(0)}}{s^{(0)}} \right] (u, \xi_0) A(du) \end{aligned}$$

Thus

$$\begin{aligned} \mathbb{U}_{n,2}(g^{(2)}) &= \mathbb{U}_{n,1}(\mathbb{E}_{\{1\}} g^{(2)}) + \mathbb{U}_{n,1}(\mathbb{E}_{\{2\}} g^{(2)}) + \mathbb{U}_{2,n}(\pi_2[g^{(2)}]) \\ &= \frac{1}{n} \sum_i \sum_h \int_0^\tau \frac{\varphi_{ih} s^{(0)} - s^{(1)}}{s^{(0)}} (u, \xi_0) M_{ih}(du) + \mathbb{U}_{2,n}(\pi_2[g^{(2)}]). \end{aligned}$$

Next, we consider the statistics $\mathbb{U}_{n,3}(g^{(3)})$. For triplets (i, j, k) of distinct indices, we have $\mathbb{E} g^{(3)}(W_i, W_j, W_k) = 0$. In addition,

$$\begin{aligned} \mathbb{E}_{\{13\}} g^{(3)}(W_i, W_j, W_k) &= \mathbb{E} g^{(3)}(W_i, W_j, W_k) | W_i, W_k \\ &= \sum_h \int_0^\tau \left[\frac{\varphi_{ih} s^{(0)} - s^{(1)}}{s^{(0)}} \frac{S_k^{(0)} - s^{(0)}}{s^{(0)}} \right] (u, \xi_0) N_{ih}(du) \\ \mathbb{E}_{\{23\}} g^{(3)}(W_i, W_j, W_k) &= \mathbb{E} g^{(3)}(W_i, W_j, W_k) | W_j, W_k \\ &= \int_0^\tau \left[\frac{s^{(1)} S_j^{(0)} - S_j^{(1)} s^{(0)}}{s^{(0)}} \left(\frac{S_k^{(0)} - s^{(0)}}{s^{(0)}} \right) \right] (u, \xi_0) A(du) \end{aligned}$$

and $\mathbb{E}_{Ag^{(3)}}(W_i, W_j, W_k) = 0$ for all other proper subsets of the index set $\{1, 2, 3\}$.

Hence

$$\mathbb{U}_{n,3}(g^{(3)}) = \mathbb{U}_{n,2}(\mathbb{E}_{\{13\}} g^{(3)}) + \mathbb{U}_{n,2}(\mathbb{E}_{\{23\}} g^{(3)}) + \mathbb{U}_{n,3}(\pi_3[g^{(3)}]).$$

Finally, for any quadruplet (i, j, k, l) of distinct indices, we have $\mathbb{E} g^{(4)}(W_i, W_j, W_k, W_l) = 0$, and

$$\mathbb{E}_{134} g^{(4)}(W_i, W_j, W_k, W_l) = \mathbb{E} [g^{(4)}(W_i, W_j, W_k, W_l) | W_i, W_k, W_l]$$

$$\begin{aligned}
&= \sum_h \int_0^\tau \frac{[|\varphi_h|s^{(0)} + \bar{s}^{(1)}]}{s^{(0)}} \left(\frac{S_k^{(0)} - s^{(0)}}{s^{(0)}}\right) \left(\frac{S_l^{(0)} - s^{(0)}}{s^{(0)}}\right) N_{hi}(du), \\
\mathbb{E}_{234}g^{(4)}(W_i, W_j, W_k, W_l) &= \mathbb{E}[g^{(4)}(W_i, W_j, W_k, W_l)|W_j, W_k, W_l] \\
&= \int_0^\tau \frac{[\bar{s}^{(1)}S_j^{(0)} + s^{(0)}\bar{S}_j^{(1)}]}{s^{(0)}} \left(\frac{S_k^{(0)} - s^{(0)}}{s^{(0)}}\right) \left(\frac{S_l - s}{s^{(0)}}\right) A(du), \\
\mathbb{E}_{34}g^{(4)}(W_i, W_j, W_k, W_l) &= \mathbb{E}[g^{(4)}(W_i, W_j, W_k, W_l)|W_k, W_l] \\
&= 2 \int_0^\tau [\bar{s}^{(1)} \left(\frac{S_k^{(0)} - s^{(0)}}{s^{(0)}}\right) \left(\frac{S_l^{(0)} - s^{(0)}}{s^{(0)}}\right)] A(du)
\end{aligned}$$

and $\mathbb{E}_A g^{(4)} = 0$ for all other proper subsets of the index set $\{1, 2, 3, 4\}$. Hence

$$\mathbb{U}_{n,4}(g^{(4)}) = \mathbb{U}_{n,2}(\mathbb{E}_{\{34\}}g^{(4)}) + \sum_{A=\{134\},\{234\}} \mathbb{U}_{n,3}(\pi_3[\mathbb{E}_A g^{(4)}]) + \mathbb{U}_{n,4}(\pi_4[g^{(4)}]).$$

In Lemma A.3, the Hoeffding decomposition of the statistics $\mathbb{U}_{n,p}(f_\xi^{(p)})$, $p = 2, 3, 4$ is quite analogous. On the other hand the remainder term is given by $\text{rem}(\xi) = \sum_{p=1}^3 \text{rem}_p(\xi)$, where $\text{rem}_1(\xi)$ is given by (4.3), and

$$\begin{aligned}
\text{rem}_2(\xi) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left[\frac{\sum_j \sum_h \varphi_{hj}(u, \xi) (S_{hj}^{(0)}(u, \xi) - S_{hj}^{(0)}(u, \xi_0))}{S^{(0)}(u, \xi_0)} \right. \\
&\quad \left. - (\xi - \xi_0)^T \frac{S^{(2)}(u, \xi_0)}{S^{(0)}(u, \xi_0)} \right] N_i(du), \\
\text{rem}_3(\xi) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left[\frac{S^{(1)}}{S^{(0)}}(u, \xi) \left(\frac{S^{(0)}(u, \xi)}{S^{(0)}(u, \xi_0)} - 1 \right) - \right. \\
&\quad \left. (\xi - \xi_0)^T \left(\frac{S^{(1)}}{S^{(0)}} \right)^{\otimes 2}(u, \xi_0) \right] N_i(du).
\end{aligned}$$

These two terms are easily verified to be order $o_p(|\xi - \xi_0|)$.