On the Smoothness of Conditional Expectation Functionals

Kyungchul Song

University of British Columbia

January 30, 2012

Abstract

Given a class Λ of real functions and a twice differentiable real-valued map φ on $\mathbf{R}$, let Γ be an $\mathbf{R}$-valued functional on Λ of form $\Gamma : \lambda \mapsto \mathbf{E}[Z \cdot \varphi(\mathbf{E}[Y|\lambda(X)])]$, where Z and Y are random variables and X is a random vector. This paper calls Γ a *conditional expectation functional*. Conditional expectation functionals often arise in semiparametric models. The main contribution of this paper is that it provides nontrivial conditions under which Γ has a uniform modulus of continuity with order 2. Hence under these conditions, the functional Γ becomes very smooth.

Keywords and Phrases: Conditional Expectation Functionals, Discretization, Uniform Modulus of Continuity, Semiparametric Models

AMS Classifications: 60, 62

1 Introduction

Let X be a random vector in $\mathbf{R}^d$ on a probability space $(\Omega, \mathcal{F}, P)$ and Λ be a collection of Borel measurable real maps on $\mathbf{R}^d$ such that for each $\lambda \in \Lambda$, $\lambda(X)$ is a continuous random variable. Endow Λ with $||\cdot||_\infty$, where $||\lambda||_\infty \equiv \sup_{x\in\mathbf{R}^d} |\lambda(x)|$, $\lambda \in \Lambda$. Let Y and Z be given random variables on $(\Omega, \mathcal{F}, P)$, and let $\varphi : \mathbf{R} \to \mathbf{R}$ be a given map. Define a functional Γ on Λ :

$$\Gamma : \lambda \longmapsto \mathbf{E}\left[Z \cdot \varphi(\mathbf{E}\left[Y|\lambda(X)\right])\right]. \tag{1}$$

This paper calls Γ a *conditional expectation functional*. Such conditional expectation functionals arise in various semiparametric models. (See Song (2011) for examples.) The main focus of this paper is on the following notion of smoothness of Γ.

Definition 1: We say that Γ has *a uniform modulus of continuity with order* α, if there exist $C > 0$ and $\varepsilon > 0$ such that for all $\eta \in (0, \varepsilon]$,

$$\sup_{\lambda_1,\lambda_2\in\Lambda:||\lambda_1-\lambda_2||_\infty\leq\eta} |\Gamma(\lambda_1) - \Gamma(\lambda_2)| \leq C\eta^\alpha.$$

Continuity of conditional expectations in the conditioning σ-fields has long received attention in the literature (e.g. Boylan (1971) and Rogge (1974).) The main departure of this paper from this literature is that the focus here is on the smoothness of a conditional expectation functional in functions indexing the conditioning variable.

The main contribution of this paper is to provide nontrivial sufficient conditions under which Γ has a uniform modulus of continuity with order $\alpha = 2$. Therefore, under these conditions, the functional Γ becomes very smooth. An important implication of this result for inference in semiparametric models is extensively studied in Song (2011).

2 Main Results

We introduce some notation. For $p \geq 1$ and a random variable W on $(\Omega, \mathcal{F}, P)$, we write $||W||_p = \{\mathbf{E}[|W|^p]\}^{1/p}$, and given a sub σ-field $\mathcal{G}$ of $\mathcal{F}$, let $\mathbf{E}[W|\mathcal{G}]$ denote a version of the conditional expectation of W given $\mathcal{G}$. We denote by $\mathcal{B}(\mathbf{R}^d)$ the Borel σ-field of $\mathbf{R}^d$. Let $\sigma(X)$ be the σ-field of X, and $\sigma(\lambda(X))$, $\lambda \in \Lambda$, be the σ-field of $\lambda(X)$. For $W \in \{Y, Z\}$ and any $A \in \mathcal{B}(\mathbf{R}^d)$ such that $P\{X \in A\} > 0$, we define $\mathbf{E}[W|X \in A] = \mathbf{E}[W1\{X \in A\}]/P\{X \in A\}$.

Let $X = [X_1^\top, X_2^\top]^\top \in \mathbf{R}^{d_1+d_2}$, where X_1 is a continuous random vector and X_2 is a discrete random vector taking values from $\{x_1, \cdots, x_M\}$. The support of a random vector is defined to be the smallest closed set in which the random vector takes values with probability one. Let S_m and $S_{1,m}$ be the supports of $X1\{X_2 = x_m\}$ and $X_1 1\{X_2 = x_m\}$, and S_X and $S_{\mathbf{E}[Y|X]}$ the supports of X and $\mathbf{E}[Y|X]$. For any set $A \subset \mathbf{R}^d$, we define $\text{diam}(A) \equiv \sup_{a,b \in A} ||a - b||$, where $|| \cdot ||$ is the Euclidean norm in $\mathbf{R}^d$. Let the Hausdorff metric between subsets A_1 and A_2 in $\mathbf{R}$ be defined by

$$d(A_1, A_2) \equiv \max \left\{ \sup_{a \in A_1} \inf_{b \in A_2} ||a - b||, \sup_{b \in A_2} \inf_{a \in A_1} ||a - b|| \right\}.$$

Also let $\lambda^{-1}(B) \equiv \{x \in \mathbf{R}^d : \lambda(x) \in B\}$ for $B \subset \mathbf{R}$.

We make the following assumptions.

Assumption 1: (i) For each $\lambda \in \Lambda$, $P\{\lambda(X) \in [0, 1]\} = 1$ and $\lambda(X)$ is a continuous random variable with a density function $f_\lambda(\cdot)$ that is bounded uniformly over $\lambda \in \Lambda$.
(ii) $\mathbf{E}[Z|X] \in [-1, 1]$ and $\mathbf{E}[Y|X] \in [-1, 1]$.

Specifying the set $[0, 1]$ in Assumption 1(i) does not cause loss of generality, as long as there exists a common bounded set in which $\lambda(X)$ takes values with probability one for all $\lambda \in \Lambda$. In such a case, the results of this paper hold only with different constants. Assumption 1(ii) requires that the conditional expectations are bounded. The number 1 expresses a scale normalization.

Assumption 2: (i) For each $m = 1, \cdots, M$, there exists $C_{1,m} > 0$ such that for each $W \in \{Y, Z\}$ and for all $s_1, s_2 \in S_{1,m}$,

$$|\mathbf{E}[W|X_1 = s_1, X_2 = x_m] - \mathbf{E}[W|X_1 = s_2, X_2 = x_m]| \leq C_{1,m}||s_1 - s_2||.$$

(ii) For each $m = 1, \cdots, M$, there exists $C_{2,m} > 0$ such that for all $\lambda \in \Lambda$, for all intervals $B_1,\ B_2 \subset \mathbf{R}$ of a finite length with $B_1 \subset B_2$ and $\lambda^{-1}(B_1) \cap S_m \neq \varnothing$, it is satisfied that

$$d(\lambda^{-1}(B_1) \cap S_m, \lambda^{-1}(B_2) \cap S_m) \leq C_{2,m} d(B_1, B_2).$$

(iii) φ is twice differentiable on the interior of $S_{\mathbf{E}[Y|X]}$ and the derivatives are bounded on $S_{\mathbf{E}[Y|X]}$.

The following lemma is an immediate consequence of Assumption 2(i).

Lemma 1: *Suppose that Assumption 2(i) holds. Then for each $m = 1, \cdots, M$, for each $W \in \{Y, Z\}$, and for any compact sets $A_1, A_2 \subset \mathbf{R}^{d_1}$ satisfying that $A_1 \subset A_2 \subset S_{1,m}$ and $P\{X_1 \in A_1, X_2 = x_m\} > 0$,*

$$|\mathbf{E}[W|X_1 \in A_2, X_2 = x_m] - \mathbf{E}[W|X_1 \in A_1, X_2 = x_m]| \leq 2C_{1,m} d(A_1, A_2),$$

where $C_{1,m} > 0$ is the constant in Assumption 2(i).

Proof: Fix $m = 1, \cdots, M$ and $A_1, A_2 \subset \mathbf{R}^{d_1}$ as in the lemma. Let $P_{j,m}$ denote the conditional distribution of X_1 given $X_1 \in A_j$ and $X_2 = x_m$, for $j \in \{1, 2\}$. Since A_1 is nonempty, fix $x^* \in A_1$ and write $\mathbf{E}[W|X_1 \in A_1, X_2 = x_m] - \mathbf{E}[W|X_1 \in A_2, X_2 = x_m]$ as

$$\int_{A_1} \left(\mathbf{E}\left[W|X_1 = x, X_2 = x_m\right] - \mathbf{E}\left[W|X_1 = x^*, X_2 = x_m\right]\right) dP_{1,m}(x)$$
$$- \int_{A_2} \left(\mathbf{E}\left[W|X_1 = x, X_2 = x_m\right] - \mathbf{E}\left[W|X_1 = x^*, X_2 = x_m\right]\right) dP_{2,m}(x),$$

because $\int_{A_1} dP_{1,m}(x) = \int_{A_2} dP_{2,m}(x) = 1$. Furthermore for $j \in \{1, 2\}$,

$$\begin{aligned} & \int_{A_j} \left| \mathbf{E}\left[W|X_1 = x, X_2 = x_m\right] - \mathbf{E}\left[W|X_1 = x^*, X_2 = x_m\right]\right| dP_{j,m}(x) \\ \leq & \sup_{s_1 \in A_j} \left| \mathbf{E}\left[W|X_1 = s_1, X_2 = x_m\right] - \mathbf{E}\left[W|X_1 = x^*, X_2 = x_m\right]\right|. \end{aligned}$$

By Assumption 2(i) and the fact that $A_1 \subset A_2$, taking the infimum over $x^* \in A_1$ on the last term yields that

$$|\mathbf{E}[W|X_1 \in A_2, X_2 = x_m] - \mathbf{E}[W|X_1 \in A_1, X_2 = x_m]| \leq 2C_{1,m} \inf_{x^* \in A_1} \sup_{s_1 \in A_2} \|s_1 - x^*\|.$$

For given $s_1 \in A_2$ and any $t < \sup_{z \in A_1} ||s_1 - z||$, the set $\{z \in A_1 : \|s_1 - z\| \leq t\}$ is compact. Furthermore, since A_2 is compact, for any $x^* \in A_1$, there exists $z^* \in A_2$ such that $\sup_{s_1 \in A_2} ||s_1 - x^*|| = ||z^* - x^*||$. Using these facts, we apply Lemma A.3 of Puhalskii and Spokoiny (1998) to conclude that

$$\inf_{x^* \in A_1} \sup_{s_1 \in A_2} \|s_1 - x^*\| = \sup_{s_1 \in A_2} \inf_{x^* \in A_1} \|s_1 - x^*\| = d(A_1, A_2).$$

Hence $|\mathbf{E}[W|X_1 \in A_2, X_2 = x_m] - \mathbf{E}[W|X_1 \in A_1, X_2 = x_m]| \leq 2C_{1,m} d(A_1, A_2)$. ■

Consider the example below where Assumption 2(ii) is satisfied.

EXAMPLE 1: Fix $\gamma_0 \in \mathbf{R}^d \backslash \{0\}$ and let $B_0 \equiv \{\gamma \in \mathbf{R}^d \backslash \{0\} : ||\gamma_0 - \gamma|| \leq \varepsilon\}$ for some $\varepsilon > 0$. Assume that there exists $\varepsilon > 0$ such that for all $\gamma \in B_0$, $X^\top \gamma$ is continuous and the density function $f_\gamma(\cdot)$ of $X^\top \gamma$ satisfies that

$$\inf_{\gamma \in B_0} \inf_{x \in S_X} f_\gamma(x^\top \gamma) > c > 0, \tag{2}$$

for some $c \in (0, 1]$, and that for each $1 \leq m \leq M$, the set $\{x^\top \gamma : x \in S_m\}$ is an interval of a finite length. For each $\gamma \in B_0$, let $\lambda_\gamma(x) \equiv F_\gamma(x^\top \gamma)$, $x \in \mathbf{R}^d$, where F_γ is the CDF of $X^\top \gamma$. Take $\Lambda \equiv \{\lambda_\gamma : \gamma \in B_0\}$. Then the class Λ satisfies Assumption 2(ii).

To see this, for each $1 \le m \le M$, we define its boundaries: $b_{L,\gamma,m} \equiv \inf\{x^\top\gamma : x \in S_m\}$ and $b_{U,\gamma,m} \equiv \sup\{x^\top\gamma : x \in S_m\}$. We define $F_\gamma^{-1}(a) \equiv \inf\{v \in \mathbf{R} : F_\gamma(v) \ge a\}$ and recall that $F_\gamma^{-1}(F_\gamma(b)) \le b$ and $F_\gamma(F_\gamma^{-1}(a)) \ge a$, for all $a, b \in \mathbf{R}$. Choose $a_1 \le a_2 \le a_3 \le a_4$, and let $B_1 \equiv [a_2, a_3]$ and $B_2 \equiv [a_1, a_4]$, such that $\lambda_\gamma^{-1}(B_1) \cap S_m \ne \varnothing$. Let $A_{1,\gamma,m} \equiv \lambda_\gamma^{-1}(B_1) \cap S_m$ and $A_{2,\gamma,m} \equiv \lambda_\gamma^{-1}(B_2) \cap S_m$, so that

$$
\begin{aligned}
A_{1,\gamma,m} &= \{x \in S_m : c_{2,\gamma,m} \le x^\top\gamma \le c_{3,\gamma,m}\}, \text{ and} \\
A_{2,\gamma,m} &= \{x \in S_m : c_{1,\gamma,m} \le x^\top\gamma \le c_{4,\gamma,m}\},
\end{aligned}
$$

where

$$
\begin{aligned}
c_{1,\gamma,m} &\equiv F_\gamma^{-1}(a_1) \vee b_{L,\gamma,m},\ c_{2,\gamma,m} \equiv F_\gamma^{-1}(a_2) \vee b_{L,\gamma,m}, \\
c_{3,\gamma,m} &\equiv F_\gamma^{-1}(a_3) \wedge b_{U,\gamma,m},\ c_{4,\gamma,m} \equiv F_\gamma^{-1}(a_4) \wedge b_{U,\gamma,m}.
\end{aligned}
$$

Note that $d(A_{1,\gamma,m}, A_{2,\gamma,m}) \le d(D_{1,\gamma,m}, D_{2,\gamma,m}) \vee d(D_{3,\gamma,m}, D_{4,\gamma,m})$, where for $j = 1, \cdots, 4$, $D_{j,\gamma,m} \equiv \{x \in S_m : c_{j,\gamma,m} = x^\top\gamma\}$. Since $d(D_{1,\gamma,m}, D_{2,\gamma,m})$ and $d(D_{3,\gamma,m}, D_{4,\gamma,m})$ are the Hausdorff distance between two parallel hyperplanes,

$$
d(D_{1,\gamma,m}, D_{2,\gamma,m}) \vee d(D_{3,\gamma,m}, D_{4,\gamma,m}) \le |c_{1,\gamma,m} - c_{2,\gamma,m}| \vee |c_{3,\gamma,m} - c_{4,\gamma,m}|.
$$

Now, if $F_\gamma^{-1}(a_1) \le b_{L,\gamma,m}$ and $F_\gamma^{-1}(a_2) \le b_{L,\gamma,m}$, $|c_{1,\gamma,m} - c_{2,\gamma,m}| = 0 \le \frac{1}{c}|a_1 - a_2|$. If $F_\gamma^{-1}(a_1) \le b_{L,\gamma,m} < F_\gamma^{-1}(a_2) < b_{U,\gamma,m}$, we have $a_1 \le F_\gamma(F_\gamma^{-1}(a_1)) \le F_\gamma(b_{L,\gamma,m}) \le a_2$, because $F_\gamma(F_\gamma^{-1}(a_2)) = F_\gamma^{-1}(F_\gamma(a_2)) = a_2$ in this case. Hence

$$
|c_{1,\gamma,m} - c_{2,\gamma,m}| \le F_\gamma^{-1}(a_2) - b_{L,\gamma,m} \le F_\gamma^{-1}(a_2) - F_\gamma^{-1}(F_\gamma(b_{L,\gamma,m})).
$$

Using (2), we deduce that

$$F_\gamma^{-1}(a_2) - F_\gamma^{-1}(F_\gamma(b_{L,\gamma,m})) \leq \frac{1}{c}\left(a_2 - F_\gamma(b_{L,\gamma,m})\right) \leq \frac{1}{c}(a_2 - a_1).$$

If $b_{L,\gamma,m} < F_\gamma^{-1}(a_1) \leq F_\gamma^{-1}(a_2) < b_{U,\gamma,m}$, then from (2), $|c_{1,\gamma,m} - c_{2,\gamma,m}| \leq F_\gamma^{-1}(a_2) - F_\gamma^{-1}(a_1) \leq \frac{1}{c}(a_2 - a_1)$. Finally, if $F_\gamma^{-1}(a_2) = b_{U,\gamma,m} = F_\gamma^{-1}(a_1)$, then $|c_{1,\gamma,m} - c_{2,\gamma,m}| = 0 \leq \frac{1}{c}|a_1 - a_2|$. Therefore, we conclude that $|c_{1,\gamma,m} - c_{2,\gamma,m}| \leq \frac{1}{c}\left|a_2 - a_1\right|$. Applying similar arguments to $|c_{3,\gamma,m} - c_{4,\gamma,m}|$, we obtain the bound $\left|a_4 - a_3\right|/c$. Hence

$$d(A_{1,\gamma,m}, A_{2,\gamma,m}) \leq \frac{1}{c}\left(|a_1 - a_2| \vee |a_3 - a_4|\right) = \frac{1}{c}d(B_1, B_2),$$

which confirms that Assumption 2(ii) is satisfied. ■

The following simple lemma forms the basis for our main result. For two sub σ-fields, $\mathcal{G}_1$ and $\mathcal{G}_2$ of $\sigma(X)$, and for a random variable W on $(\Omega, \mathcal{F}, P)$, define

$$\delta_W(\mathcal{G}_1, \mathcal{G}_2) \equiv \left\|\mathbf{E}\left[W|\mathcal{G}_1, X_2\right] - \mathbf{E}\left[W|\mathcal{G}_2, X_2\right]\right\|_2.$$

LEMMA 2: *Suppose that Assumptions 1(ii) and 2(iii) are satisfied. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be sub σ-fields of $\sigma(X)$ such that $\mathcal{G}_2 \subset \mathcal{G}_1$. Then*

$$\begin{aligned} & \left|\mathbf{E}\left[Z\{\varphi(\mathbf{E}[Y|\mathcal{G}_1]) - \varphi(\mathbf{E}[Y|\mathcal{G}_2])\}\right]\right| \\ \leq \; & (\bar{\varphi}_1 + (\bar{\varphi}_2/2)) \cdot \{\delta_Y(\mathcal{G}_1, \mathcal{G}_2) \cdot \delta_Z(\mathcal{G}_1, \mathcal{G}_2) + \delta_Y^2(\mathcal{G}_1, \mathcal{G}_2)\}, \end{aligned}$$

where $\bar{\varphi}_1 \equiv \sup_{y \in S_{\mathbf{E}[Y|X]}} |\varphi'(y)|$ and $\bar{\varphi}_2 \equiv \sup_{y \in S_{\mathbf{E}[Y|X]}} |\varphi''(y)|$.

PROOF: By using Assumptions 1(ii) and 2(iii), and expanding $\varphi(\mathbf{E}[Y|\mathcal{G}_1]) - \varphi(\mathbf{E}[Y|\mathcal{G}_2])$, we bound $\left|\mathbf{E}\left[Z\{\varphi(\mathbf{E}[Y|\mathcal{G}_1]) - \varphi(\mathbf{E}[Y|\mathcal{G}_2])\}\right]\right|$ by

$$\left|\mathbf{E}\left[Z\varphi'(\mathbf{E}[Y|\mathcal{G}_2])\{\mathbf{E}[Y|\mathcal{G}_1] - \mathbf{E}[Y|\mathcal{G}_2]\}\right]\right| + (\bar{\varphi}_2/2)\mathbf{E}\left[\{\mathbf{E}[Y|\mathcal{G}_1] - \mathbf{E}[Y|\mathcal{G}_2]\}^2\right].$$

The second term is bounded by $(\bar{\varphi}_2/2)\delta_Y^2(\mathcal{G}_1,\mathcal{G}_2)$. By the law of iterated conditional expectation, the first term is written as

$$\begin{aligned} & |\mathbf{E}\left[\{\mathbf{E}[Z|\mathcal{G}_1]-\mathbf{E}[Z|\mathcal{G}_2]\}\varphi'(\mathbf{E}[Y|\mathcal{G}_2])\{\mathbf{E}[Y|\mathcal{G}_1]-\mathbf{E}[Y|\mathcal{G}_2]\}\right]| \\ \leq\ & \bar{\varphi}_1\cdot\delta_Z(\mathcal{G}_1,\mathcal{G}_2)\cdot\delta_Y(\mathcal{G}_1,\mathcal{G}_2), \end{aligned}$$

yielding the desired result. ■

We introduce a discretization of $\lambda(X)$. Fix $\Delta\in(0,1/2]$ and define $J_\Delta\equiv\lfloor 1/\Delta\rfloor$, so that $1/\Delta-1<J_\Delta\leq 1/\Delta$, where $\lfloor a\rfloor$, $a\in\mathbf{R}$, denotes a greatest integer that is smaller than or equal to a. For each $j=1,\cdots,J_\Delta-1$, $\mathcal{I}_{j,\Delta}\equiv[(j-1)\Delta,j\Delta)$, and for $j=J_\Delta$, $\mathcal{I}_{j,\Delta}\equiv[(J_\Delta-1)\Delta,1]$, so that $\mathrm{diam}(\mathcal{I}_{j,\Delta})=\Delta$ for $j=1,\cdots,J_\Delta-1$, and $\Delta\leq\mathrm{diam}(\mathcal{I}_{j,\Delta})$ $<2\Delta$ for $j=J_\Delta$. Then let

$$\lambda^\Delta(X)\equiv\sum_{j=1}^{J_\Delta}a_{j,\Delta}\cdot 1\left\{\lambda(X)\in\mathcal{I}_{j,\Delta}\right\}, \tag{3}$$

where $a_{j,\Delta}\equiv(j-1)\Delta$.

Lemma 3: *Suppose that Assumptions 1(i) and 2(i)(ii) hold. Then for each $W\in\{Y,Z\}$ and for all $\Delta\in(0,1/2]$,*

$$\mathrm{sup}_{\lambda\in\Lambda}\delta_W(\sigma(\lambda(X)),\sigma(\lambda^\Delta(X)))\leq 4C_{1,2}\Delta,$$

where $C_{1,2}\equiv\max_{1\leq m\leq M}C_{1,m}C_{2,m}$ and $C_{1,m}$ and $C_{2,m}$ are the constants in Assumptions 2(i) and (ii).

Proof: Fix $W\in\{Y,Z\}$, $\Delta\in(0,1/2]$, $\Delta_0\in(0,\Delta]$, and $\lambda\in\Lambda$. For each $j=1,\cdots,J_\Delta$, we partition $\mathcal{I}_{j,\Delta}=\cup_{k=1}^{K_{j,\Delta}}\mathcal{J}_{jk,\Delta}$ into $K_{j,\Delta}$ intervals of $\mathcal{J}_{jk,\Delta}$ with a positive length in $[\Delta_0,2\Delta_0)$. Let $\pi_m\equiv P\{X_2=x_m\}$. Define $q_{jk,m}\equiv P\{\lambda(X)\in\mathcal{J}_{jk,\Delta}|X_2=x_m\}$, $q_{j,m}\equiv P\{\lambda(X)\in$

$\mathcal{I}_{j,\Delta}|X_2 = x_m\}$, whenever $\pi_m > 0$. First, we define a discretized version of λ as follows:

$$\tilde{\lambda}(X) \equiv \sum_{j=1}^{J_\Delta} \sum_{k=1}^{K_{j,\Delta}} b_{jk,\Delta} \cdot 1\left\{\lambda(X) \in \mathcal{J}_{jk,\Delta}\right\},$$

where $b_{jk,\Delta}$ is the left-end point of the interval $\mathcal{J}_{jk,\Delta}$. Then, certainly, $\lambda^\Delta(X) = \tilde{\lambda}^\Delta(X)$, because $\tilde{\lambda}(X)$ is a finer discretization of $\lambda(X)$ than $\lambda^\Delta(X)$ is. It suffices to show that when we replace $\lambda(\cdot)$ with $\tilde{\lambda}(\cdot)$, we can obtain the same bound in Lemma 3 that does not depend on the choice of Δ_0. Let $\mathbb{L}_{j,m} \equiv \{1 \leq k \leq K_{j,\Delta} : P\{\lambda(X) \in \mathcal{J}_{jk,\Delta}, X_2 = x_m\} > 0\}$ and bound $\delta_W^2(\sigma(\tilde{\lambda}(X)), \sigma(\tilde{\lambda}^\Delta(X)))$ by

$$\begin{aligned} & \sum_{m=1}^{M} \sum_{j=1}^{J_\Delta} \sum_{k\in\mathbb{L}_{j,m}} \left\{ R_{jk,m} - \sum_{r\in\mathbb{L}_{j,m}} R_{jr,m} \frac{q_{jr,m}}{q_{j,m}} \right\}^2 q_{jk,m}\pi_m \qquad (4) \\ \leq \; & \sum_{m=1}^{M} \sum_{j=1}^{J_\Delta} \sum_{k\in\mathbb{L}_{j,m}} \sum_{r\in\mathbb{L}_{j,m}} \left\{R_{jk,m} - R_{jr,m}\right\}^2 \frac{q_{jr,m}}{q_{j,m}} q_{jk,m}\pi_m, \end{aligned}$$

where $R_{jk,m} = \mathbf{E}\left[W|\lambda(X) \in \mathcal{J}_{jk,\Delta}, X_2 = x_m\right]$. By Lemma 1 and Assumption 2(ii), we find that

$$\begin{aligned} |R_{jk,m} - R_{jr,m}| &\leq 2C_{1,m} d\left(\lambda^{-1}(\mathcal{J}_{jk,\Delta}) \cap S_m, \lambda^{-1}(\mathcal{J}_{jr,\Delta}) \cap S_m\right) \\ &\leq 2C_{1,m}C_{2,m} d\left(\mathcal{J}_{jk,\Delta}, \mathcal{J}_{jr,\Delta}\right) < 4C_{1,m}C_{2,m}\Delta, \end{aligned}$$

because both $\mathcal{J}_{jk,\Delta}$ and $\mathcal{J}_{jr,\Delta}$ are sub-intervals of $\mathcal{I}_{j,\Delta}$ and the length of $\mathcal{I}_{j,\Delta}$ is bounded by 2Δ. By plugging this bound into the last sum in (4), we obtain the wanted result. ■

Lemmas 2 and 3 can be used to control the difference $|\Gamma(\lambda) - \Gamma(\lambda^\Delta)|$ uniformly over $\lambda \in \Lambda$: for all $\Delta \in (0, 1/2]$,

$$\sup_{\lambda\in\Lambda} |\Gamma(\lambda) - \Gamma(\lambda^\Delta)| \leq 16C_{1,2}^2 \left\{2\bar{\varphi}_1 + \bar{\varphi}_2\right\} \Delta^2. \qquad (5)$$

It remains to find a bound for $|\Gamma(\lambda_1^\Delta) - \Gamma(\lambda_2^\Delta)|$ that is uniform over $\lambda_1, \lambda_2 \in \Lambda$ with λ_1 and λ_2 close to each other. This is fulfilled by Lemma 4 below. The proof of Lemma 4 is lengthy and found in the appendix.

LEMMA 4: *Suppose that Assumptions 1-2 hold. Then for all* $\Delta \in (0, 1/2]$, $q \in (0, \Delta/16)$, *and for all* $\lambda_1, \lambda_2 \in \Lambda$ *such that* $||\lambda_1 - \lambda_2||_\infty \leq q$,

$$|\Gamma(\lambda_1^\Delta) - \Gamma(\lambda_2^\Delta)| \leq C_3 \Delta q,$$

where $C_3 \equiv (9 + 192\bar{c}_f)\{2\bar{\varphi}_1 + \bar{\varphi}_1\}C_{1,2}^2$, $\bar{c}_f \equiv \sup_{\lambda \in \Lambda} \sup_{x \in S_X} f_\lambda(\lambda(x))$, *and* $C_{1,2} > 0$ *is the constant in Lemma 3.*

With some $a > 16$ and $\Delta = aq$ in Lemma 4, the bound becomes $C_3 a q^2$. Hence, by a proper choice of discretization, one can control substantially the order of maximal deviation of $\Gamma(\lambda^\Delta)$ at the local perturbation of λ. Now, we are prepared to prove the main result of this paper.

THEOREM 1: *Suppose that Assumptions 1-2 hold. Then* Γ *has a uniform modulus of continuity with order* 2.

PROOF: Fix $\varepsilon > 0$ and $q \in (0, \varepsilon]$ and take $\Delta \in (16q, 17q]$. Choose $\lambda_1, \lambda_2 \in \Lambda$ such that $||\lambda_1 - \lambda_2||_\infty \leq q$. Bound $|\Gamma(\lambda_1) - \Gamma(\lambda_2)|$ by

$$\begin{aligned} & \left|\Gamma(\lambda_1) - \Gamma(\lambda_1^\Delta)\right| + \left|\Gamma(\lambda_1^\Delta) - \Gamma(\lambda_2^\Delta)\right| + \left|\Gamma(\lambda_2^\Delta) - \Gamma(\lambda_2)\right| \\ \leq \;& 2C_4\Delta^2 + C_3\Delta q \leq (2C_4 + C_3)q^2(17^2 + 17), \end{aligned}$$

where $C_4 \equiv 16C_{1,2}^2\{2\bar{\varphi}_1 + \bar{\varphi}_2\}$. The first inequality follows by (5) and Lemma 4. ■

3 Appendix: Proof of Lemma 4

For any $\lambda_2 \in \Lambda$ and $q > 0$, let $\lambda_{2,q}(x) \equiv \lambda_2(x)+2q$ and $\lambda_{2,q}^{\Delta}(x) \equiv (\lambda_2(x)+2q)^{\Delta}$, the discretized version of $\lambda_2(x) + 2q$. Also define for $W \in \{Y, Z\}$, $\lambda_1, \lambda_2 \in \Lambda$, $q > 0$ and $\Delta > 0$,

$$\begin{aligned} e_W(\lambda_1^{\Delta}, \lambda_{2,q}^{\Delta}) &\equiv \mathbf{E}\left[W|\lambda_1^{\Delta}(X), \lambda_{2,q}^{\Delta}(X), X_2\right] - \mathbf{E}\left[W|\lambda_1^{\Delta}(X), X_2\right] \text{ and} \\ e_W(\lambda_{2,q}^{\Delta}, \lambda_1^{\Delta}) &\equiv \mathbf{E}\left[W|\lambda_1^{\Delta}(X), \lambda_{2,q}^{\Delta}(X), X_2\right] - \mathbf{E}\left[W|\lambda_{2,q}^{\Delta}(X), X_2\right]. \end{aligned}$$

Lemma A1: *Suppose that Assumptions 1(i) and 2(i)(ii) hold. Then for any* $q \in (0, 1/2]$, $\Delta > 16q$, *for* $W \in \{Y, Z\}$, *and any* $\lambda_1, \lambda_2 \in \Lambda$ *such that* $||\lambda_1 - \lambda_2||_\infty \leq q$,

$$\mathbf{E}\left[e_W^2(\lambda_1^{\Delta}, \lambda_{2,q}^{\Delta})\right] \leq C_5 \Delta q, \ \textit{and } \mathbf{E}\left[e_W^2(\lambda_{2,q}^{\Delta}, \lambda_1^{\Delta})\right] \leq C_5 \Delta q,$$

where $C_5 \equiv ((9/4) + 48\bar{c}_f)C_{1,2}^2$ *and* $C_{1,2} > 0$ *is the constant in Lemma 3.*

Proof: We only show that $\mathbf{E}[e_W^2(\lambda_1^{\Delta}, \lambda_{2,q}^{\Delta})] \leq C_5 \Delta q$. Let $\overline{\mathcal{I}}_{j,\Delta}$ be the closure of $\mathcal{I}_{j,\Delta}$. By Assumption 1(ii), for any $j = 1, \cdots, J_\Delta$,

$$\left|\mathbf{E}\left[W 1\{\lambda(X) \in \overline{\mathcal{I}}_{j,\Delta} \backslash \mathcal{I}_{j,\Delta}\}\right]\right| \leq P\{\lambda(X) \in \overline{\mathcal{I}}_{j,\Delta} \backslash \mathcal{I}_{j,\Delta}\} = 0, \text{ for all } \lambda \in \Lambda,$$

because $\lambda(X)$ is continuous. Therefore, replacing $\mathcal{I}_{j,\Delta}$'s by $\overline{\mathcal{I}}_{j,\Delta}$'s does not alter $\mathbf{E}[e_W^2(\lambda_1^{\Delta}, \lambda_{2,q}^{\Delta})]$. For $j, s = 1, \cdots, J_\Delta$, let $p_{js,m} \equiv P\{\lambda_1(X) \in \overline{\mathcal{I}}_{j,\Delta}, \lambda_{2,q}(X) \in \overline{\mathcal{I}}_{s,\Delta}, X_2 = x_m\}$ and

$$\begin{aligned} \overline{\mathcal{I}}_{j,\Delta}^{-\alpha} &\equiv \{a \in \overline{\mathcal{I}}_{j,\Delta} : a + \varepsilon \in \overline{\mathcal{I}}_{j,\Delta} \text{ for all } \varepsilon \in [-\alpha, \alpha]\} \text{ and} \\ \overline{\mathcal{I}}_{j,\Delta}^{\alpha} &\equiv \{a + \varepsilon \in \mathbf{R} : a \in \overline{\mathcal{I}}_{j,\Delta}, \ \varepsilon \in [-\alpha, \alpha]\}. \end{aligned}$$

Define $\mathbb{N}_\Delta \equiv \{1, \cdots, J_\Delta\}$, and $\mathbb{N}_{1,m} \equiv \{(s, j) \in \mathbb{N}_\Delta^2 : p_{js,m} > 0\}$. We first collect some facts.

Fact 1: For all $m = 1, \cdots, M$, and for all $s \neq j$, $p_{js,m} \leq 3\bar{c}_f q$.

Proof: Note that $p_{js,m} \leq P\{\lambda_1(X) \in \overline{\mathcal{I}}_{j,\Delta}, \lambda_{2,q}(X) \in \overline{\mathcal{I}}_{s,\Delta}\} \leq P\{\lambda_1(X) \in \overline{\mathcal{I}}_{j,\Delta} \cap \overline{\mathcal{I}}_{s,\Delta}^{3q}\}$.

Fact 1 follows because the last probability is bounded by $\bar{c}_f\mu(\overline{\mathcal{I}}_{j,\Delta}\cap\overline{\mathcal{I}}^{3q}_{s,\Delta})\leq 3\bar{c}_f q$, where, for $A\in\mathcal{B}(\mathbf{R})$, $\mu(A)$ denotes Lebesgue measure of A.

Fact 2: For all $m=1,\cdots,M$, and all $s\neq j$ with $(s,j)\in\mathbb{N}_{1,m}$, $s=j+1$.

Proof: Note that $3q\geq\inf_{x\in S_X}\{\lambda_{2,q}(x)-\lambda_1(x)\}=\inf_{x\in S_X}\{\lambda_2(x)-\lambda_1(x)\}+2q\geq q$, and that $16q<\Delta\leq$diam$(\mathcal{I}_{j,\Delta})$ for all $j=1,\cdots,J_\Delta$. Since $||\lambda_1-\lambda_2||_\infty\leq q$, $p_{js,m}>0$ and $s\neq j$, it is necessarily that $s=j+1$.

Fact 3: For all $m=1,\cdots,M$, and all $s\neq j$ with $(s,j)\in\mathbb{N}_{1,m}$,

$$\begin{aligned}\{\lambda_1(X) &\in \overline{\mathcal{I}}_{j,\Delta}\cap\overline{\mathcal{I}}^{q}_{s,\Delta}\}\subset\{\lambda_1(X)\in\bar{I}_{j,\Delta},\text{ and }\lambda_{2,q}(X)\in\bar{I}_{s,\Delta}\}\\ &\subset \{\lambda_1(X)\in\overline{\mathcal{I}}_{j,\Delta}\cap\overline{\mathcal{I}}^{3q}_{s,\Delta}\}.\end{aligned}$$

Proof: The second inclusion is straightforward. As for the first inclusion, since $||\lambda_1-\lambda_2||_\infty\leq q$ and $s=j+1$ by Fact 2, if $\lambda_1(X)\in\overline{\mathcal{I}}_{j,\Delta}\cap\overline{\mathcal{I}}^{q}_{s,\Delta}$, then $\lambda_{2,q}(X)=\lambda_2(X)+2q\in\overline{\mathcal{I}}_{s,\Delta}$.

Now we are prepared to prove Lemma A1. For $j=1,\cdots,J_\Delta$, let

$$E_{1j,m}\equiv\{\lambda_1(X)\in\overline{\mathcal{I}}_{j,\Delta},X_2=x_m\},\;\; E_{2j,m}\equiv\{\lambda_{2,q}(X)\in\overline{\mathcal{I}}_{j,\Delta},X_2=x_m\}$$

and for $\alpha>0$, let $E^{[-\alpha]}_{2j,m}\equiv\{\lambda_{2,q}(X)\in\overline{\mathcal{I}}^{-\alpha}_{j,\Delta},X_2=x_m\}$ and $E^{[\alpha]}_{2j,m}\equiv\{\lambda_{2,q}(X)\in\overline{\mathcal{I}}^{\alpha}_{j,\Delta},X_2=x_m\}$. Note that

$$\begin{aligned}\mathbf{E}\left[e^2_W(\lambda^\Delta_1,\lambda^\Delta_{2,q})\right] &= \sum_{m=1}^M\mathbf{E}\left[e^2_W(\lambda^\Delta_1,\lambda^\Delta_{2,q})|X_2=x_m\right]\pi_m\\ &\leq \sum_{m=1}^M\sum_{(s,j)\in\mathbb{N}_{1,m}:j=s}\left(\mathbf{E}(W|E_{1j,m}\cap E_{2s,m})-\mathbf{E}(W|E_{1j,m})\right)^2p_{js,m}\pi_m\\ &\quad+\sum_{m=1}^M\sum_{(s,j)\in\mathbb{N}_{1,m}:j\neq s}\left(\mathbf{E}(W|E_{1j,m}\cap E_{2s,m})-\mathbf{E}(W|E_{1j,m})\right)^2p_{js,m}\pi_m.\end{aligned}$$

Since $p_{js,m}=P(E_{1j,m}\cap E_{2s,m})>0$ for all $(s,j)\in\mathbb{N}_{1,m}$, we apply Lemma 1 to bound

$(\mathbf{E}(W|E_{1j,m} \cap E_{2s,m}) - \mathbf{E}(W|E_{1j,m}))^2$ by

$$4C_{1,m}^2 d^2\left(\lambda_1^{-1}(\overline{\mathcal{I}}_{j,\Delta}) \cap \lambda_{2,q}^{-1}(\overline{\mathcal{I}}_{s,\Delta}) \cap S_m, \lambda_1^{-1}(\overline{\mathcal{I}}_{j,\Delta}) \cap S_m\right).$$

Now, if $j = s$,

$$\begin{aligned} & d^2\left(\lambda_1^{-1}(\overline{\mathcal{I}}_{j,\Delta}) \cap \lambda_{2,q}^{-1}(\overline{\mathcal{I}}_{s,\Delta}) \cap S_m, \lambda_1^{-1}(\overline{\mathcal{I}}_{j,\Delta}) \cap S_m\right) \\ \leq \; & d^2\left(\lambda_1^{-1}(\overline{\mathcal{I}}_{j,\Delta} \cap \overline{\mathcal{I}}_{j,\Delta}^{-3q}) \cap S_m, \lambda_1^{-1}(\overline{\mathcal{I}}_{j,\Delta}) \cap S_m\right) \leq C_{2,m}^2 d^2\left(\overline{\mathcal{I}}_{j,\Delta}^{-3q}, \overline{\mathcal{I}}_{j,\Delta}\right) \leq 9C_{2,m}^2 q^2, \end{aligned}$$

by Assumption 2(ii). On the other hand, if $j \neq s$, then by Facts 2 and 3,

$$\begin{aligned} & d^2\left(\lambda_1^{-1}(\overline{\mathcal{I}}_{j,\Delta}) \cap \lambda_{2,q}^{-1}(\overline{\mathcal{I}}_{s,\Delta}) \cap S_m, \lambda_1^{-1}(\overline{\mathcal{I}}_{j,\Delta}) \cap S_m\right) \\ \leq \; & d^2\left(\lambda_1^{-1}(\overline{\mathcal{I}}_{j,\Delta} \cap \overline{\mathcal{I}}_{s,\Delta}^{q}) \cap S_m, \lambda_1^{-1}(\overline{\mathcal{I}}_{j,\Delta}) \cap S_m\right) \leq C_{2,m}^2 d^2\left(\overline{\mathcal{I}}_{j,\Delta} \cap \overline{\mathcal{I}}_{j+1,\Delta}^{q}, \overline{\mathcal{I}}_{j,\Delta}\right) \\ \leq \; & C_{2,m}^2 \mathrm{diam}^2\left(\overline{\mathcal{I}}_{j,\Delta}\right) \leq 4C_{2,m}^2 \Delta^2. \end{aligned}$$

Therefore, we bound $\mathbf{E}\left[e_W^2(\lambda_1^{\Delta}, \lambda_{2,q}^{\Delta})\right]$ by

$$36C_{1,2}^2 q^2 \sum_{m=1}^{M} \sum_{(s,j)\in\mathbb{N}_{1,m}:j=s} p_{js,m}\pi_m + 16C_{1,2}^2 \Delta^2 \sum_{m=1}^{M} \sum_{(s,j)\in\mathbb{N}_{1,m}:j\neq s} p_{js,m}\pi_m.$$

The first term is bounded by $36C_{1,2}^2 q^2$ because $\sum_{m=1}^{M} \sum_{(s,j)\in\mathbb{N}_{1,m}:j=s} p_{js,m}\pi_m \leq 1$ and the second term is bounded by $48C_{1,2}^2 \Delta^2 J_\Delta \bar{c}_f q \leq 48C_{1,2}^2 \bar{c}_f \Delta q$, because the number of the pairs (s,j) in $\mathbb{N}_{1,m}$ such that $j \neq s$ is bounded by J_Δ (by Fact 2), $J_\Delta \leq 1/\Delta$, and $p_{js,m} \leq 3\bar{c}_f q$ by Fact 1. Therefore, we conclude that

$$\mathbf{E}\left[e_W^2(\lambda_1^{\Delta}, \lambda_{2,q}^{\Delta})\right] < (9/4)C_{1,2}^2 \Delta q + 48C_{1,2}^2 \bar{c}_f \Delta q = ((9/4) + 48\bar{c}_f)C_{1,2}^2 \Delta q,$$

where the first inequality follows because $\Delta > 16q$. ■

Proof of Lemma 4: With $\lambda^{\Delta}_{2,q}(X)$ defined prior to Lemma A1, write $\Gamma(\lambda^{\Delta}_1) - \Gamma(\lambda^{\Delta}_{2,q})$ as

$$\begin{aligned} &\mathbf{E}\left[Z\{\varphi(\mathbf{E}[Y|\lambda^{\Delta}_1(X)]) - \varphi(\mathbf{E}[Y|\lambda^{\Delta}_1(X), \lambda^{\Delta}_{2,q}(X)])\}\right] \\ &+\mathbf{E}\left[Z\{\varphi(\mathbf{E}[Y|\lambda^{\Delta}_1(X), \lambda^{\Delta}_{2,q}(X)]) - \varphi(\mathbf{E}[Y|\lambda^{\Delta}_{2,q}(X)])\}\right]. \end{aligned} \tag{6}$$

We use Lemma 2 to bound the absolute value of the first expectation by

$$\left(\bar{\varphi}_1 + \frac{\bar{\varphi}_2}{2}\right) \cdot ||e_Y(\lambda^{\Delta}_1, \lambda^{\Delta}_{2,q})||_2 \cdot \left\{||e_Z(\lambda^{\Delta}_1, \lambda^{\Delta}_{2,q})||_2 + ||e_Y(\lambda^{\Delta}_1, \lambda^{\Delta}_{2,q})||_2\right\}. \tag{7}$$

The above is bounded by $\{2\bar{\varphi}_1 + \bar{\varphi}_2\}C_5\Delta q$ by Lemma A1. We use the same argument for the second term in (6) to deduce that $|\Gamma(\lambda^{\Delta}_1) - \Gamma(\lambda^{\Delta}_{2,q})| \leq 2\{2\bar{\varphi}_1 + \bar{\varphi}_2\}C_5\Delta q$. Applying this inequality two times, we bound $|\Gamma(\lambda^{\Delta}_1) - \Gamma(\lambda^{\Delta}_2)|$ by

$$|\Gamma(\lambda^{\Delta}_1) - \Gamma(\lambda^{\Delta}_{2,q})| + |\Gamma(\lambda^{\Delta}_{2,q}) - \Gamma(\lambda^{\Delta}_2)| \leq 4\{2\bar{\varphi}_1 + \bar{\varphi}_2\}C_5\Delta q,$$

yielding the desired bound. ■

4 Acknowledgement

The author is grateful to the referee, whose comments have led to the improved presentation of the paper.

References

[1] Boylan, E. S. (1971), "Equi-convergence of martingales," *Annals of Mathematical Statistics*, 42, 552-559.

[2] Puhalskii, A. and V. Spokoiny (1998): "On large-deviation efficiency in statistical inference," *Bernoulli*, 4, 203-272.

[3] Rogge, L. (1974): "Uniform inequalities for conditional expectations," *Annals of Probability*, 2, 486-489.

[4] Song, K. (2011): "Semiparametric models with single-index nuisance parameters," Working Paper, University of British Columbia.