ANNÉE 2015

UNIVERSITÉ DE RENNES 1

ueb

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Mathématiques et applications

Ecole doctorale Matisse

présentée par

Weiyu LI

préparée à l'IRMAR (UMR CNRS 6625) et au CREST
Institut de recherche mathématiques de Rennes
Centre de recherche en économie et statistique
École Nationale de la Statistique et de l'Analyse de l'Information

---

Quelques contributions à l'estimation des modèles définis par des équations estimantes conditionnelles

Thèse soutenue à l'Ensai
le 15 juillet 2015

devant le jury composé de :

Jean-Yves Dauxois
Professeur INSA de Toulouse / rapporteur

Cédric Heuchenne
Professeur University of Liège / rapporteur

Laura Dumitrescu
Associate Research Professor ENSAI / examinateur

Olivier Lopez
Maître de Conférences Université Pierre et Marie Curie / examinateur

Laurent Rouviere
Maître de Conférences Université Rennes 2 / examinateur

Valentin Patilea
Professeur ENSAI / directeur de thèse

Acknowledgements

In the last four years, I had a very wonderful time in Rennes. I got married and finished this thesis. During this period, I met many kindly persons and got a lot of help from them. Foremost, I would like to express my appreciation to Professor PATILEA Valentin for offering me the opportunity of studying in France. Thanks him for his guidance, support and help. He is endlessly kind and patient with me. I'm deeply grateful to him for everything he did for me. I also want to thank my dissertation committee: Professor DAUXOIS Jean-Yves, Professor HEUCHENNE Cédric, Doctor DUMITRESCU Laura, Doctor LOPEZ Olivier and Doctor ROUVIERE Laurent. Thanks them for their time and insights on the thesis.

Sincere thanks are given to all friends in France. They helped me and brought me joy. I would especially like to thank YOU Rui and her boy friend BRACH Loïc, they are so kindly and friendly to offer help whenever I need. Thanks to MAISTRE Samuel who helped me a lot with my work. Thanks to all my colleagues in Ensai and the secretaries of Matisse and IRMAR. Thanks to the CSC secretaries ZHAO Jingmei and WU Xueming who are generous in giving help. I would like to thank my family for the irreplaceable roles they have played. Thanks to my parents for their love and unreserved support in all my pursuits. Thanks to my husband Weizhi for helping me to apply for the scholarship of CSC, adjusting to the life in France and accompanying me through these days.

Finally, I would like to send my gratitude and blessing to my motherland, who is funding thousands of youth like me chasing their dreams overseas.

Contents

1

Introduction

Une grande partie de modèles statistiques et économétriques (régressions, régressions quantiles, modèles de transformation, modèles à variables instrumentales, *etc.*), peuvent se définir sous la forme des équations de moments conditionnels, encore appélées *équations estimantes conditionnelles*. Plus précisément, soit $Z = (Y^\top, X^\top)^\top$ un vecteur aléatoire dont nous supposons observer des réalisations indépendantes. (Pour toute matrice A, nous notons par $A^\top$ sa transposée. Dans ce manuscrit, tous les vecteurs sont des matrices colonne.) Soit $g(Z, \beta, \eta_\beta)$ une fonction mesurable donnée, dépendant d'un paramètre de dimension finie β, appartenant à un ensemble de paramètres $\mathcal{B}$, et d'un paramètre de dimension infinie η, qui a son tour pourrait dépendre de β et de Z. La fonction g peut être à valeurs scalaires ou vectorielles. Le paramètre d'intérêt est β, le paramètre η est souvent appelé *paramètre de nuisance*. Une forme générale d'un modèle défini par des équations estimantes conditionnelles est la suivante : il existe un unique $\beta_0 \in \mathcal{B}$ tel que

$$\mathbb{E}\left[g(Z, \beta_0, \eta_{\beta_0}(Z)) \mid X\right] = 0, \quad \text{presque sûrement.} \tag{.1}$$

Voir Ai & Chen (2003) pour une étude de ce modèle général.

Dans cette thèse nous allons considérer d'abord le cas où g ne dépend pas d'une fonction inconnue η_β et le modèle peut se réécrire sous une forme plus simple : il existe un unique $\beta_0 \in \mathcal{B}$ tel que

$$\mathbb{E}\left[g(Z, \beta_0) \mid X\right] = 0, \quad \text{presque sûrement.} \tag{.2}$$

Ensuite, nous allons etudier le cas où g dépend d'une telle fonction η_β qui se définit sous la forme d'une régression. Plus précisément, nous allons nous concentrer sur les modèles semi paramétriques dites à direction révélatrice unique (*single-index* en anglais).

Le modèle (.2), très étudié dans la littérature économétrique, est lui-même un modèle assez général. En particulier, il inclut :

1. les modèles de régression paramétriques classiques, linéaire ou non linéaire (notamment les modèles linéaires généralisés, comme la régression logistique, le modèle probit, la régression poissonnienne, *etc.*), dans quel cas

$$g(Z, \beta) = Y - \mu(Z, \beta),$$

où $\{\mu(\cdot, \beta) : \beta \in \mathcal{B}\}$ est le modèle de régression considéré;

2. les modèles de régression quantile, en définissant

$$g(Z, \beta) = \mathbf{1}\{Y \leq q(Z, \beta)\} - \rho,$$

pour un $\rho \in (0, 1)$ fixé et $\{q(\cdot, \beta) : \beta \in \mathcal{B}\}$ un modèle de régression pour le quantile conditionnel d'ordre ρ, par exemple un modèle de régression quantile linéaire (nous notons par $\mathbf{1}\{\cdot\}$ la fonction indicatrice);

3. les modèles de régression avec transformation de la variable à expliquer, en considérant

$$g(Z, \beta) = h(Y, \lambda) - r(X, \theta),$$

où $\{h(\cdot, \lambda) : \lambda \in \Lambda\}$ est une famille de transformations, par exemple la famille de transformations Box-Cox, et $\{r(\cdot, \theta) : \theta \in \Theta\}$ est un modèle de régréssion pour la variable transformée;

4. le modèle d'évaluation des actifs basé sur la consommation (*CCAPM* ou *consumption-based capital asset pricing model* en anglais) de Hansen & Singleton (1983), ainsi qu'une large gamme des modèles économétriques;

5. les modèles de régression en présence de l'endogénéité pour lesquels on procède

à une approche par variables instrumentales, dans quel cas on peut définir

$$g(Z, \beta) = \tilde{Y} - m(W, \beta),$$

avec $\{m(\cdot, \beta) : \beta \in \mathcal{B}\}$ le modèle de régression, le vecteur W des régresseurs endogènes, X un vecteur de variables instrumentales et $Z = (Y^\top, X^\top)^\top$ avec $Y = (\tilde{Y}, W^\top)^\top$.

Il est également possible d'avoir plusieurs équations en même temps, ce qui correspondrait à une fonction q à valeurs vectorielles. Cela peut être le cas, par exemple, lorsqu'on souhaite modéliser simultanément l'espérance conditionnelle et la variance conditionnelle. Parmi d'autres nombreuses contributions dans la littérature, Kitamura *et al.* (2004), Diminguez & Lobato (2004), Qin & Lawless (2004) présentent plusieurs exemples de modélisation amenant à des équations estimantes comme dans l'équation (.2).

Une approche habituelle pour procéder à l'inférence statistique à partir des équations (.2) consiste à réécrire ces équations sous forme des moments non conditionnels. La raison est donnée par le fait qu'un moment non conditionnel peut être facilement approché par une moyenne empirique. Par conséquent, une équation estimante basée sur l'échantillon observé peut s'écrire et se résoudre, produisant ainsi un estimateur. Les outils probabilistes (loi de grands nombres, théorème central limite, *etc*) permettent alors de contrôler l'erreur d'approximation de l'équation estimante et en déduire le comportement des estimateurs construits à partir de l'approximation. Cela est bien plus difficile avec une équation estimante conditionnelle car, en général, on ne dispose pas d'un nombre suffisant d'observations pour chaque valeur de la variable conditionnante. En général, une procédure non paramétrique qui cherche de l'information autour de chaque valeur de la variable conditionnante est alors nécessaire.

Le passage des équations estimantes conditionnelles vers des équations estimantes non conditionnelle se réalise habituellement par une fonction mesurable $A(X)$ à valeurs dans un espace de matrices, avec un nombre de colonnes égal à la dimension de g et un nombre de lignes, en général, égal ou supérieur à la dimension du

paramètre à estimer β. Autrement dit, en général on considère les équations

$$\mathbb{E}\left[A(X)g(Z,\beta)\right] = 0, \qquad \beta \in \mathcal{B}, \tag{.3}$$

et on suppose que β_0 est identifié par ces conditions. Bien qu'il soit facile de se convaincre que le β_0 identifié par le modèle (.1) vérifie aussi les équations (.3), il n'est pas garanti que β_0 soit la seule valeur du paramètre à vérifier ces dernières équations. En effet, en général, une équation de moment conditionnel est équivalente à un nombre dénombrable d'équations de moments non conditionels, ce qui explique la perte d'identifiabilité dans le modèle (.3). Voir Dominguez & Lobato (2004) pour quelques exemples élémentaires de non identifiabilité.

En général, considérer un nombre infini d'équations serait infaisable en pratique. Par conséquent, plusieurs solutions ont été proposées pour pallier à ce problème: faire tendre le nombre d'équations estimantes vers l'infini, comme, par exemple, dans Carrasco & Florens (2000); utiliser la vraisemblance empirique locale, voir Kitamura *et al.* (2004); utiliser une statistique de type Cramer-Von Mises qui permet de préserver l'identifiabilité, voir Dominguez & Lobato (2004). Plus récemment, Lavergne & Patilea (2013) ont proposé une approche alternative, qu'ils ont appelée SMD (*smooth minimum distance* en anglais). Cette approche peut s'interpréter comme une équation de type (.3) avec une fonction $A(X)$ bien choisie. Bien qu'inspiré d'un lissage non paramétrique par noyau, le critère d'estimation utilisé par les auteurs ne nécessite pas que le paramètre de lissage, la fenêtre, tend vers zéro. C'est cette approche introduite par Lavergne & Patilea (2013), et rappelée dans le Chapitre 2, qui représente le fil conducteur de cette thèse. Dans un premier temps, nous proposons une version itérative de l'approche SMD qui évite l'optimisation non linéaire. A chaque étape de l'itération, la solution du problème d'optimisation est explicite. Ainsi, l'approche SMD devient faisable pour des modèles avec des paramètres de grande dimension.

Ensuite, dans ce travail de thèse, nous nous intéressons au modèle plus général (.1). Plus précisément, nous considérons le problème d'estimation des modèles à direction révélatrice unique (*single-index* en anglais). Les modèles semi paramétriques

de type single index (SIM) sont maintenant largement utilisés par les statisticiens et économètres. Dans le cas de régression, l'hypothèse de SIM indique que l'espérance conditionnelle de la variable réponse Y sachant le vecteur des variables explicatives X est identique à celle sachant une combinaison linéaire des variables explicatives. Récemment, cette idée a été généralisée à la régression quantile. Lorsque l'on s'intéresse à la distribution conditionnelle d'une variable observée sachant les variables explicatives, sous l'hypothèse de SIM la variable réponse est conditionnellement indépendante des variables explicatives sachant une combinaison linéaire de ces variables. L'approche de modélisation de type SIM est une idée naturelle de réduction de la dimension qui réalise un compromis entre une modélisation purement paramétrique et celle purement non paramétrique. Comme expliqué dans le Chapitre 3, les modèles SIM que nous avons mentionné peuvent se réécrire sous la forme d'un modèle d'equations estimantes comme (.1). Par exemple, dans le cas d'une régression sous l'hypothèse SIM, on écrira

$$\mathbb{E}\left[Y - \mathbb{E}(Y \mid X^\top \beta) \mid X\right] = 0, \quad \text{presque sûrement,}$$

et on supposera qu'il existe une unique direction donnée par un vecteur β_0 qui vérifie cette équation estimante conditionnelle.

Bien qu'il existe aujourd'hui plusieurs techniques pour estimer un modéle sous une hypothèse SIM, la plupart de ces méthodes permettent uniquement d'estimer des modèles de régression. Assez peu de méthodes existent pour des modèles à distribution conditionnelle. Dans cette thèse, on propose d'étendre l'idée de SMD aux modèles SIM et ainsi construire une nouvelle approche d'estimation semi paramétrique pour ces modèles. Ceci est réalisé dans le Chapitre 3. Cette approche s'applique aussi bien à des SIM en régression que aux SIM en loi. Des comparaisons par simulations avec des méthodes existantes montrent des bonnes performances pour notre approche.

Dans la dernière partie de la thèse nous considérons les modèles SIM dans le contexte de données censurées. Dans des nombreuses applications, par exemple en biostatistique, en fiabilité, ou encore en économie du travail, on souhaite modéliser

la loi d'une durée d'intérêt à l'aide des covariables, en présence d'un mécanisme de censure. Le modèle à risques proportionnel de Cox (1972) représente l'approche la plus utilisée. Il propose néanmoins une structure particulière pour la loi conditionnelle de la durée d'intérêt, qui vérifie en particulier une condition de SIM. L'idée est alors de considérer une modélisation SIM, sans faire référence à un modèle particulier, comme celui de Cox. De telles modélisations ont apparu récemment dans le littérature, voir Bouaziz & Lopez (2010), Xia *et al.* (2010) et Strzalkowska-Kominiak & Cao (2013). Nous utilisons une idée plus simple pour écrire le modèle et ensuite nous faisons appel à l'approche SMD pour l'estimation de la direction β_0. Pour ce faire, on utilise une remarque de Dabrowska (1989) qui rappele que l'estimateur de Kaplan-Meier est une fonctionnelle lisse de la loi empirique des observations (les durées et les indicatrices de censure). Ceci reste vrai pour la version conditionnelle de l'estimateur Kaplan-Meier proposé par Beran (1981). Par conséquent, nous proposons d'imposer les conditions SIM dans l'espace des lois des observations, estimer la direction de β_0 dans cet espace à l'aide de l'approche SMD, et ensuite utiliser la fonctionnelle Kaplan-Meier conditionnelle pour estimer la loi conditonnelle de la durée d'intérêt. Ceci résulte en une méthodologie simple, qui permet, en particulier, de tester les hypothèse de réduction de la dimension.

I Le résumé du Chapitre 2

Soit $Z = (Y^\top, X^\top)^\top \in \mathbb{R}^{d+q}$ un vecteur aléatoire. Soit $\beta \in \mathcal{B} \subset \mathbb{R}^p$ un vecteur des paramètres à estimer, dans un espace de dimension finie. Considérons le modèle d'équations estimantes conditionelles suivant : il existe un unique $\beta_0 \in \mathcal{B}$,

$$\mathbb{E}[g(Z, \beta_0)|X] = 0 \quad a.s.,$$

où $g(Z, \beta_0)$ est une fonction à valeurs dans $\mathbb{R}^r$ connue, $r \geq 1$.

Au lieu de suivre l'approche classique et de transformer ces équations dans un système de dimension finie d'équations non conditionnelles, nous poursuivons l'approche proposée par Lavergne & Patilea (2013). Ainsi nous évitons les problèmes

d'identification, et donc de consistance, produit par le remplacement des équations estimantes conditionnelles par des équations estimantes non conditionnelles. L'idée de l'approche de Lavergne & Patilea se base sur le faits suivants : si $K(\cdot)$ est un noyau avec une transformée de Fourier positive sur $\mathbb{R}^q$, pour tout $h > 0$,

$$\mathbb{E}\left[g(Z_1, \beta)^\top g(Z_2, \beta) h^{-q} K((X_1 - X_2)/h)\right] \geq 0,$$

et l'égalité est vérifiée si et seulement si

$$\mathbb{E}[g(Z, \beta)|X] f(X) = 0 \quad a.s.,$$

où $f(\cdot)$ est la densité de X. La condition que X ait une densité peut se relaxer sans difficulté, mais pour une présentation plus simple nous ne remettons pas en cause cette condition.

La dernière équation montre que β_0 peut être défini comme le minimum unique de l'application

$$\beta \mapsto \mathbb{E}\left[g(Z_1, \beta)^\top g(Z_2, \beta) h^{-q} K((X_1 - X_2)/h)\right] \geq 0.$$

Dès lors un estimateur naturel de β_0 est obtenu en minimisant une approximation empirique de cette espérance. Dans cette procédure, la fenêtre h ne devra pas nécessairement tendre vers zéro pour établir les résultats asymptotiques habituels (consistance et normalité asymptotique). Plus précisement, l'estimateur de Lavergne et Patilea est définit comme un minimum du critère

$$M_{n,h}(\beta) = \frac{1}{2n(n-1)} \sum_{1 \leq i \neq j \leq n} g(Z_i, \beta)^\top g(Z_j, \beta) K_{ij}, \quad K_{ij} = \frac{1}{h^q} K\left(\frac{X_i - X_j}{h}\right).$$

L'estimateur définit comme minimum de $M_{n,h}(\beta)$ nécessite une optimisation non linéaire. Dans le premier chapitre de la thèse, nous proposons une version itérative de cet estimateur qui évite l'optimisation linéaire. Pour la construire, nous utilisons une linéarisation de $g(Z_i, \beta)$ et de $g(Z_j, \beta)$ à l'aide d'un développement de Taylor à l'ordre 1. En mettant dans les gradients la valeur du paramètre égale à la précédente valeur

dans le procédé d'itération, le problème à résoudre devient quadratique, avec une solution explicite. Aux pas suivants, on met à jour le paramètre dans les gradients, et on minimise la nouvelle forme quadratique, *etc.*

Les itérations obtenues ainsi sont similaires aux itération de type Newton-Kantorovitch (voir, par exemple, le livre de Kantorovitch & Akilov (1964)) pour trouver les racines du gradient de $M_{n,h}(\beta)$. Ici, la matrice hessienne est simplifiée en enlevant un terme qui est censé être petit si β est proche de β_0. De cette manière, la hessienne sera toujours semi-définie positive. Il reste à espérer que les itérations restent dans un voisinage suffisamment petit de β_0. Nous étudions les performances de cette méthode itérative à l'aide des nombreuses simulations. Pour ce faire, nous considérons deux modèles: un modèle logistique et un modèle de régression avec des régresseurs endogènes. Dans chaque cas, nous comparons les résultats avec les méthodes classiques. Nous considérons également un modèle de régression non linéaire et le problème de sélection des variables explicatives par pénalisation de type LASSO. Notre méthode itérative montre des bonne propriétés empiriques dans les situations considérées. Les propriétés théoriques restent à explorer dans un prolongement du travail de thèse.

II Le résumé du Chapitre 3

Dans ce chapitre, nous allons proposer une nouvelle approche d'inférence dans les modèles à direction révélatrice unique. Les modèles à direction révélatrice unique (*single-index* en anglais) représentent une idée naturelle de réduction de la dimension. Cette approche réalise un compromis entre une modélisation purement paramétrique et celle purement non paramétrique.

Soit Y un vecteur des variables à expliquer et soit X un vecteur de d variables explicatives. Soit T_u, $u \in \mathcal{U}$, une famille de transformations de Y. Dans une approche de régression à direction révélatrice unique, on suppose qu'il existe un unique β_0 tel que

$$E[T_u(Y)|X] = E[T_u(Y)|X^\top \beta_0], \qquad \forall u \in [0,1]. \tag{II.4}$$

Le vecteur β_0 est le paramètre de dimension finie à estimer, il appartient à un ensemble de paramètres

$$\mathcal{B} \subset \{(\beta_1, \dots \beta_d) : \beta_1 = 1\} \subset \mathbb{R}^d.$$

En particulier, dans ce cadre nous pouvons intégrer les deux contextes suivants.

(a) **Régression semi paramétrique à direction révélatrice unique**: dans ce cas on prend $T_u(y) = y$, $\forall u$, et par conséquent

$$E[Y|X] = E[Y|X^\top \beta_0].$$

Voir, parmi beaucoup d'autres, les articles Hall *et al.* (1993), Hristache *et al.* (2001), Delecroix *et al.* (2006), Xia *et al.* (2002), Cui *et al.* (2011).

(b) **Loi conditionnelle à direction révélatrice unique**: supposons $Y \in \mathbb{R}$ (le cas d'un Y multi dimensionnel se traite de manière très similaire) et soit $T_u(y) = \mathbf{1}\{y \leq \Phi^{-1}(u)\} = \mathbf{1}\{\Phi(y) \leq u\}$, avec $\Phi(\cdot)$ une transformation qui ramène Y sur $[0, 1]$ (par exemple $\Phi(\cdot)$ est la fonction de répartition gaussienne de même moyenne et même variance que Y); dans ce cas, la condition (II.4) signifie que la loi conditionnelle de Y sachant X est identique à la loi conditionnelle de Y sachant $X^\top \beta_0$. Parmi d'autres auteurs, ce cadre a été étudié auparavant par Delecroix *et al.* (2003), Hall & Yao (2005).

La méthode d'estimation que nous proposons se base sur l'idée suivante. Soit $f_\beta(\cdot)$ la densité de $X^\top \beta$, qui est supposée exister pour chaque $\beta \in \mathcal{B}$. Soit

$$g_u(Y, X, \beta) = \{T_u(Y) - \mathbb{E}[T_u(Y)|X^\top \beta]\} f_\beta(X^\top \beta), \qquad u \in \mathcal{U}, \;\; \beta \in \mathcal{B}.$$

Alors, la condition (II.4) se traduit par la condition

$$\mathbb{E}[g_u(Y, X, \beta)|X] = 0, \;\; \forall u \in \mathcal{U} \qquad \text{si et seulement si} \qquad \beta = \beta_0.$$

Ensuite nous appliquons l'idée de l'approche SMD avec cette fonction g.

La difficulté supplémentaire est donnée par le fait que g contient des fonctions

inconnues qu'il faut estimer par une méthode non paramétrique. Autrement dit, dans le cas d'une seule transformation T_u (comme par exemple pour la régression) cette fois-ci nous définissons l'estimateur comme le minimum de

$$\widehat{M}_{n,h}(\beta) = \frac{1}{2n(n-1)} \sum_{1 \le i \ne j \le n} \widehat{g}(Y_i, X_i, \beta)^\top \widehat{g}(Y_j, X_j, \beta) K_{ij},$$

avec $\widehat{g}$ un estimateur par noyau de g. Dans le cas d'une famille non dégénérée des transformations T_u, le critère $\widehat{M}_{n,h}(\beta)$ dépendra aussi de u, dans quel cas nous proposons d'intégrer le critère par rapport à u avant procéder à sa minimisation pour trouver l'estimateur.

Par la construction de la fonction g, l'estimateur à noyau de g ne comporte pas des dénominateurs. Cela nous permettra d'éviter l'utilisation des fonctions d'élagage (*trimming function* en anglais) et de permettre aux variables explicatives d'avoir un support non borné. Notre méthode semble être la seule à permettre ces deux aspects.

Nous montrons que l'estimateur est consistant et $\sqrt{n}-$asymptotiquement normal. Sa variance asymptotique étant compliquée, nous proposons une procédure de type bootstrap pour calculer des intervalles de confiance. Cette procédure par simulation est basée sur une idée de perturbation aléatoire du critère $\widehat{M}_{n,h}(\beta)$: construire

$$\widehat{M}^*_{n,h}(\beta) = \frac{1}{2n(n-1)} \sum_{1 \le i \ne j \le n} \widehat{g}(Y_i, X_i, \beta)^\top \widehat{g}(Y_j, X_j, \beta) K^*_{ij},$$

avec $K^*_{ij} = K_{ij}\xi_i\xi_j$, où $\xi_1, \cdots, \xi_n$ est un échantillon i.i.d., indépendant de l'échantillon observé, des variables exponentielles de paramètre égal à 1; minimiser $\widehat{M}^*_{n,h}(\beta)$ pour obtenir un estimateur de type bootstrap; répéter l'opération plusieurs fois de manière indépendante et utiliser les estimateurs de type bootstrap pour déduire la loi de l'estimateur initial.

Nous présenterons les résultats des plusieurs expériences par simulation et en utilisant des données réelles pour évaluer la performance de notre nouvel estimateur. Les résultats empiriques sont encourageants.

III Le résumé du Chapitre 4

Dans le dernier chapitre nous considérons l'estimation de la fonction de répartition conditionnelle en présence de censure. Nous proposons une idée originale de réduction de la dimension, en utilisant une hypothèse de type single-index. Ensuite, nous estimons la direction β à l'aide d'un approche de type SMD.

Soit T une variable aléatoire à valeurs dans $(-\infty, \infty]$. Dans les modèles de durées on suppose souvent que T est supérieure ou égale à 0, mais formellement nous n'avons pas besoin de cette contrainte. Considérons le cas où la statisticienne observe un échantillon indépendant des variables Y, δ et X, avec Y réelle, δ une variable indicateur et X un vecteur de régresseurs à valeurs dans un espace $\mathcal{X}$. La variable indicateur nous dit si Y est précisément égal à T ou juste plus petit que T. Autrement dit,

$$\delta = 1 \quad \text{if} \quad Y = T \qquad \text{and} \qquad \delta = 0 \quad \text{if} \quad Y < T.$$

L'objectif est d'estimer la loi de T sachant X. Afin de rendre notre modélisation à toute une catégorie d'applications en biostatistique (voir les *cure models*), économie de travail (voir le *split population model*), finance et assurance, nous permettrons que la probabilité conditionnelle de l'évènement $\{T = \infty\}$ soit positive.

Les observations sont caractérisées par les sous-probabilités conditionnelles

$$\begin{aligned} H_1((-\infty, t] \mid x) &= \mathbb{P}(Y \leq t, \delta = 1 \mid X = x) \\ H_0((-\infty, t] \mid x) &= \mathbb{P}(Y \leq t, \delta = 0 \mid X = x), \qquad t \in \mathbb{R},\ x \in \mathcal{X}. \end{aligned}$$

La loi de Y est caractérisée par

$$H((-\infty, t] \mid x) = \mathbb{P}(Y \leq t \mid X = x) = H_0((-\infty, t] \mid x) + H_1((-\infty, t] \mid x).$$

La manière usuelle de modéliser cette situation est de supposer qu'il existe une

variable C, le temps de censure, et que

$$Y = T \wedge C, \qquad \delta = \mathbf{1}\{T \leq C\}.$$

Sous des hypothèses d'identification appropriées, par exemple, T et C sont indepéndentes sachant X, il est possible d'exprimer sous une forme explicite la fonction de répartition, ou la survie, de T sachant X en fonction de $H_0(\cdot \mid x)$ et $H_1(\cdot \mid x)$. En insérant des estimateur de $H_0(\cdot \mid x)$ et $H_1(\cdot \mid x)$ dans cette expression explicite, on obtient un estimateur pour la loi conditionnelle de T. Lorsqu'on estime H_0 et H_1 par les fonctions de répartition empiriques (sans tenir compte de X), on retrouve l'estimateur Kaplan-Meier de la loi de T. Lorsqu'on estime H_0 et H_1 par un estimateur de type Nadaraya-Watson, on retrouve l'estimateur de Beran (1981); voir aussi Dabrowska (1989).

Afin d'éviter le *fléau de la dimension* lorsqu'on dispose de plusieurs variables explicatives, dans le Chapitre 4 nous proposons une technique de réduction de la dimension de type single-index, comme celle étudiée dans le Chapitre 3. L'originalité vient du fait que cette condition est imposé sur les variables observées (Y, δ). Plus precisément, en supposant $\mathcal{X} = \mathbb{R}^d$, nous imposons que

$$(Y, \delta) \perp X \mid X^\top \beta_0,$$

pour un vecteur $\beta_0 \in \mathcal{B} \subset \mathbb{R}^d$ à estimer. Pour estimer β_0, nous adaptons l'approche proposée dans le Chapitre 3 au cas de données censurées. Nous prouvons des résultats asymptotiques de convergence en probabilité et normalité asymptotique. Nous proposons des intervalles de confiance par la technique de perturbation aléatoire présentée dans le Chapitre 3. Enfin, nous proposons un estimateur de type single-index pour la probabilité conditionnelle de l'évènement $\{T = \infty\}$, généralisant ainsi les résultats de Xu & Peng (2014).

Il est important de mentionner que, contrairement aux approches existantes de modélisation de type single-index en présence de censure, voir Bouaziz & Lopez (2010), Xia *et al.* (2010) et Strzalkowska-Kominiak & Cao (2013), dans notre approche l'hypothèse single-index peut se tester facilement, en utilisant l'approche de

Maistre & Patilea (2014).

Nous présenterons les résultats des plusieurs expériences par simulation et en utilisant des données réelles pour évaluer la performance de la nouvelle approche.

IV References

1. Ai, C., and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* **71**, 1795-1843.

2. Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley.

3. Bouaziz, O. & Lopez, O. (2010). Conditional density estimation in a censored single-index model. *Bernoulli* **16**, 514–542.

4. Carrasco, M., and Florens, J.P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory* **16**, 797-834.

5. Cox, D.R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

6. Cui X., Härdle W., & Zhu L. (2011). The EFM approach for single-index models. *Ann. Statist.* **39**, 1658-1688.

7. Dabrowska, D. M. (1989). Uniform Consistency of the Kernel Conditional Kaplan-Meier Estimate. *Ann. Statist.* **17**, 1157–1167.

8. Delecroix, M., Härdle, W. & Hristache, M. (2003). Efficient estimation in conditional single-index regression. *J. Multivariate Anal.* **86**, 213–226.

9. Delecroix, M., Hristache, M. & Patilea, V. (2006). On semiparametric $M-$estimation in single-index regression. *J. Statist. Plan. Inference* **136**, 730–769.

10. Dominguez, M.A., and Lobato, I.N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica* **72**, 1601-1615.

11. Hall, P. & Yao, Q. (2005). Approximating conditional distribution functions using dimension reduction. *Ann. Statist.* **33**, 1404–1421.

12. Härdle W., Hall P. & Ichimura H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21**, 157–178.

13. Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometric* **50**, 1029-1054.

14. Hansen, L.P., and Singleton, K.J. (1983). Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns. *J. Political Economy* **91**, 249-265.

15. Horowitz, J. (2009). *Semiparametric and Nonparametric Methods in Econometrics.* Springer Series in Statistics. Springer: New-York.

16. Hristache M., Juditsky A. & Spokoiny V. (2001). Direct estimation of the index coefficient in a single-index model. *Ann. Statist.* **29**, 595–917.

17. Kantorovich, L.V., and Akilov, G.P. (1964), *Functional Analysis in Normed Spaces*. Pergamon Press.

18. Kitamura, Y., Tripathi, G., and Ahn, H. (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* **72**, 1667-1714.

19. Lavergne, P., and Patilea, V. (2013). Smooth minimum distance estimation and testing in conditional moment restrictions models: uniform in bandwidth theory. *J. Econometrics* **177**, 47-59.

20. Lee, K.-Y., Li, B. & Chiaromonte, F. (2013). A general theory for nonlinear sufficient dimension reduction: Formulation and estimation. *Ann. Statist.* **41**, 221–249.

21. Maistre, S. & Patilea, V. (2014). Nonparametric model checks of single-index assumptions. arXiv:1410.4934 [math.ST].

22. Qin, J., and Lawless, J. (2004). Empirical likelihood and generalized estimating equations. *Ann. Statist.* **22**, 300-325.

23. Strzalkowska-Kominiak, E. & Cao, R. (2013). Maximum likelihood estimation for conditional distribution single-index models under censoring. *J. Multivariate Anal.* **114**, 74-98.

24. Xia, Y., Tong, H., Li, W. K. & Zhu, L. (2002) An adaptive estimation of dimension reduction space (with discussion). *J Royal Statist. Soc. Series B* **64**, 363-410.

25. Xia, Y., Dixin Zhang, D. & Xu, J. (2010). Dimension Reduction and Semiparametric Estimation of Survival Models. *J Amer. Statist. Assoc.* **105(489)**, 278-290.

26. Xu, J., & Peng, Y. (2014). Nonparametric cure rate estimation with covariates. *Canad.J. Statist.* **42**, 1–17.

27. Zheng, D., Yin, G. & Ibrahim, J.G. (2006). Semiparametric Transformation Models for Survival Data With a Cure Fraction. *J. Amer. Statist. Assoc.* **101(474)**, 670-684.

2

Iterative Estimation for Conditional Estimating Equations

I Introduction

Let $Z = (Y^\top, X^\top)^\top \in \mathbb{R}^{d+p}$ be a random vector and $\beta \in \mathcal{B} \subset \mathbb{R}^p$ be a unknown vector of parameters in some finite dimensional parameter set. A Conditional Estimating Equation (CEE) model is defined as follows: for some unique $\beta_0 \in \mathcal{B}$,

$$\mathbb{E}[g(Z, \beta_0)|X] = 0 \quad a.s., \tag{I.1}$$

where $g(Z, \beta_0)$ is a known r-vector valued function with $r \geq 1$. The CEE models are quite general, many common models can fit into it, such as the mean regression models, where $g(Z, \beta) = Y - \mu(X, \beta)$, the conditional quantile models, where $g(Z, \beta) = \mathbf{1}\{Y - q(X, \beta) \leq 0\} - \rho$ for a quantile of order $\rho \in (0, 1)$, the mean regression models with a nonlinear transformation of the dependent variable, e.g. the Box-Cox transformation, where $g(Z, \theta, \lambda) = h(Y, \lambda) - X^\top\theta$, the instrumental variables regressions, the econometric models of optimizing agents, *etc.*

The most common methods to estimate the finite dimensional parameter β_0 identified by the CEE (I.1) proceed, at a preliminary stage, to the transformation of the CEE into some unconditional estimating equations

$$\mathbb{E}[A(X)g(Z, \beta)] = 0 \qquad a.s., \beta \in \mathcal{B}, \tag{I.2}$$

where $A(X)$, sometimes called the *instruments* matrix, is a matrix valued function of X chosen by the user. In general, the matrix $A(X)$ has at least as many lines as the dimension of β and may depend on the unknown β_0. Next, one could apply any method for unconditional moment equations, as for instance the Generalized Method of Moments (GMM) of Hansen (1982) or the Empirical Likelihood (EL) of Qin & Lawless (1994). By a suitable choice of the matrix $A(X)$ one could achieve the semiparametric efficiency bound for the estimation of β_0. See Newey (1993).

It is important to keep in mind that the equations (I.2) may not identify the parameter β_0. In other words, several values of β could satisfy the equations (I.2). See, for instance, Dominguez & Lobato (2004) for some simple examples of this identification failure. As a consequence, any estimation method based on the equations (I.2) could fail to yield consistent estimators. As a remedy of this aspect, several procedures have been proposed. Some methods rely on increasing the number of lines in the matrix $A(X)$ (in other words, increasing the number of instruments) with the sample size, such as the minimum distance approach of Ai & Chen (2003), or generalizations of GMM and EL by Donald *et al.* (2003) and Hjort *et al.* (2009). Carrasco & Florens (2000) generalize the GMM approach to a continuum of estimating equations. Other EL-type estimators use nonparametric smoothing to estimate conditional equations, such as Antoine *et al.* (2007), Kitamura *et al.* (2004), and Smith (2007a,b). All these approaches involve a use-chosen parameters (number of estimating equations, regularization parameter, or smoothing parameter) that could have large influence on the estimation result. Dominguez & Lobato (2004) proposed a new method based on Cramer-von-Misses criterion which does not require any user-chosen parameter, but it is in general not semiparametrically efficient.

Lavergne & Patilea (2013) propose a new class of smooth minimum distance (SMD) estimators based on quadratic contrasts built by kernel smoothing. They develop a theory that focuses on uniformity in bandwidth and established a $\sqrt{n}-$asymptotic representation of they estimator as a process indexed by a bandwidth. The bandwidth can vary within a wide range that includes bandwidths independent of the sample size. The theoretical result hold true even when the model (I.1) is misspecified. In the case of a well specified model, efficient estimators could be obtained

by suitable weighting of the quadratic contrast.

Here, we reconsider the SMD method proposed of Lavergne & Patilea (2013) and we propose an iterative approach for building the estimates. The new idea is based on a linearization of the function $g(z;\beta)$ and results in an iterative quadratic minimization procedure with explicit solution at each stage. Thus the iterative version of the SMD approach avoids nonlinear optimization and this could be very useful when applying this approach with high-dimensional parameters.

The chapter is organized as follows. In section II we recall the SMD approach, while in section III we introduce our iterative version. In particular, we show that the iterative approach could be easily extended to a penalized version of the SMD criterion and thus could serve for variable selection. Section IV presents several simulation studies that reveals that our approach performs well in applications.

II The Smooth Minimum Distance approach

Let $\{Z_1, \ldots, Z_n\}$ be an independent sample from Z. The Smooth Minimum Distance (SMD) estimator, introduced by Lavergne & Patilea (2013) for inference in CEE models as in equation (I.1), is defined as a solution of the following optimization problem:

$$\tilde{\beta}_{n,h} = \arg\min_{\mathcal{B}} M_{n,h}(\beta) \tag{II.1}$$

where

$$M_{n,h}(\beta) = \frac{1}{2n(n-1)} \sum_{1\le i\ne j\le n} g(Z_i,\beta)^\top g(Z_j,\beta) K_{ij}, \quad K_{ij} = \frac{1}{h^q} K\left(\frac{X_i - X_j}{h}\right) \tag{II.2}$$

with $K(\cdot)$ a multivariate kernel and $h = h_n$ a sequence of bandwidths. Here and in the following, for a matrix A, $A^\top$ stands for its transpose.

Let $\mathcal{F}[K](u) = (2\pi)^{-q/2} \int_{\mathbb{R}^q} \exp(-iu^\top t) K(t) dt$, $u \in \mathbb{R}^q$, denote the Fourier Transform of $K(\cdot)$. To understand why the minimum of the criterion $M_{n,h}(\beta)$ could be a consistent estimator, it suffices to compute the expectation of $M_{n,h}(\beta)$. Let $g_k(\cdot,\beta)$, $1 \le k \le r$, denote the components of the vector values function $g(\cdot,\beta)$. For simplicity, assume that X has a density and let $f(\cdot)$ be that density. However, as pointed

out by Lavergne & Patilea (2013), the SMD approach could be easily accommodated to take into account discrete variables, if X has such components. By the Inverse Fourier Transform, Fubini Theorem and the independence of the observations,

$$\begin{aligned}
\mathbb{E}M_{n,h}(\beta) &= \frac{1}{2}\,\mathbb{E}\left[g(Z_1,\beta)^\top g(Z_2,\beta)h^{-q}K((X_1-X_2)/h)\right] \\
&= \frac{1}{2}\,(2\pi)^{-q/2}\mathbb{E}\left[g(Z_1,\beta)^\top g(Z_2,\beta)\int_{\mathbb{R}^q}\exp(it^\top(X_1-X_2))\mathcal{F}[K](ht)dt\right] \\
&= \frac{1}{2}\,(2\pi)^{q/2}\sum_{k=1}^{r}\left\{\int_{\mathbb{R}^q}\left|\mathcal{F}\left[\mathbb{E}[g_k(Z,\beta)|X=\cdot\,]\,f(\cdot)\right](t)\right|^2\mathcal{F}[K](ht)dt\right\} \\
&\geq 0, \qquad \text{(II.3)}
\end{aligned}$$

where the last inequality holds provided $\mathcal{F}[K](\cdot)$ is strictly positive on $\mathbb{R}^q$. Moreover,

$$\begin{aligned}
\mathbb{E}M_{n,h}(\beta)=0 &\Leftrightarrow \mathcal{F}[\mathbb{E}[g_k(Z,\beta)|X=\cdot]f(\cdot)](t)=0 \;\; \forall t\in\mathbb{R}^q,\; k=1,\ldots,r \\
&\Leftrightarrow \mathbb{E}[g(Z,\beta)|X]f(X)=0 \;\; a.s. \\
&\Leftrightarrow \beta=\beta_0.
\end{aligned}$$

In other words, even for a fixed bandwidth h, the limit contrast $\mathbb{E}M_{n,h}(\beta)$ identifies the parameter β_0 defined by the CEE model (I.1). As a consequence, the SMD estimator $\tilde{\beta}_{n,h}$ is expected to consistently estimated β_0, even with a fixed bandwidth h. The assumption on the positiveness of the Fourier Transform of the kernel $K(\cdot)$ is mild and fulfilled, for instance, by products of normal, logistic, Laplace, and Student densities.

Lavergne & Patilea (2013) derived the consistency and an i.i.d. asymptotic representation of $\tilde{\beta}_{n,h}$, uniformly with respect to h in a range including constant bandwidths. The results follow from the asymptotic behavior of the $M_{n,h}(\beta)$. Semi-parametric efficiency could be achieved by replacing in the definition of $M_{n,h}(\beta)$ the vector valued function $g(Z,\beta)$ by the weighted vector valued function $\{\mathrm{Var}[g(Z,\beta_0)\mid X]\}^{-1/2}g(Z,\beta)f^{-1/2}(X)$. Here and in the following, for a positive definite matrix A, $A^{1/2}$ denote its square root, that is a positive definite matrix with the square equal to A. Moreover, $A^{-1/2}$ denote the inverse of the square root of A. Since in general $\mathrm{Var}[g(Z,\beta_0)\mid X]f(X)$ depends on the unknown β_0, efficiency could be achieved

through a two-step procedure: first compute $\tilde{\beta}_{n,h}$ the minimum of $M_{n,h}(\beta)$ defined in equation (II.2); next, estimate $\mathrm{Var}[g(Z,\tilde{\beta}_{n,h}) \mid X]f(X)$ nonparametrically and minimize $M_{n,h}(\beta)$ redefined with the estimate of $\{\mathrm{Var}[g(Z,\tilde{\beta}_{n,h}) \mid X]\}^{-1/2} g(Z,\beta) f^{-1/2}(X)$ instead of $g(Z,\beta)$. A prominent example where $\mathrm{Var}[g(Z,\tilde{\beta}_{n,h}) \mid X]$ does not depend on β_0 is provided by the quantile regressions. Indeed, in that case $r = 1$ and $\mathrm{Var}[g(Z,\tilde{\beta}_{n,h}) \mid X] = \rho(1-\rho)$, so that one could achieve efficiency in one-step after replacing $f(\cdot)$ by a nonparametric estimator and thus building an estimate of the weighted functions $\{\rho(1-\rho)\}^{-1/2} g(Z,\beta) f^{-1/2}(X)$. Lavergne & Patilea (2013) also noticed that their asymptotic results remain valid even if one includes the diagonal terms, those corresponding to equal indices i and j, in the definition of $M_{n,h}(\beta)$. To define our iterative version of the SMD estimator, we will include the diagonal terms in the definition of $M_{n,h}(\beta)$.

To improve the asymptotic approximations of the law of the SMD estimators, in particular to account for the effect of the bandwidth h, one could build a bootstrap version of the SMD estimators. For this purpose, consider $\xi_1, \cdots, \xi_n$ an i.i.d. sample of exponential random variables with parameter equal to 1, independent of the sample of Z. Then, let

$$\tilde{\beta}^*_{n,h} = \arg\min_{\mathcal{B}} M^*_{n,h}(\beta),$$

where

$$M^*_{n,h}(\beta) = \frac{1}{2n(n-1)} \sum_{1 \le i \ne j \le n} g(Z_i,\beta)^\top g(Z_j,\beta) K^*_{ij},$$

and

$$K^*_{ij} = K_{ij}\xi_i\xi_j, \qquad 1 \le i,j \le n. \tag{II.4}$$

Lavergne & Patilea (2013) showed that conditionally of the sample of the observations, the asymptotic law of $\sqrt{n}(\tilde{\beta}^*_{n,h} - \tilde{\beta}_{n,h})$ is centered normal with the same variance as $\sqrt{n}(\tilde{\beta}_{n,h} - \beta_0)$. From such a result, one could derive, for instance, confidence intervals for the components of β_0.

Let us close this section with an interpretation of the SMD approach which could explain why one may expect good performance with small samples. For this purpose, let us consider the case of mean regression with $g(Z,\beta) = Y - r(Z,\beta)$, where Y is

some univariate response and $\{r(\cdot,\beta) : \beta \in \mathcal{B}\}$ is a parametric mean regression model. Then we can write

$$\begin{aligned} 2\mathbb{E}M_{n,h}(\beta) &= \mathbb{E}\left[\mathbb{E}\{g(Z_1,\beta) \mid X_1\}\mathbb{E}\{g(Z_2,\beta) \mid X_2\}h^{-q}K((X_1-X_2)/h)\right] \\ &= \mathbb{E}\left[\{\mathbb{E}(Y_1 \mid X_1) - r(X_1,\beta)\}\{\mathbb{E}(Y_2 \mid X_2) - r(Y_2,\beta)\}h^{-q}K((X_1-X_2)/h)\right], \end{aligned}$$

and the last expectation tends to

$$\mathbb{E}\left[\{\mathbb{E}(Y \mid X) - r(X,\beta)\}^2 f(X)\right],$$

as h tends to zero. This shows that in the case of a linear regression model, the SMD approach is closely related to class of estimators introduced by Cristóbal-Cristóbal *et al.* (1987) and based on the presmoothing of the response variable. Faraldo-Roca & González-Manteiga (1987) showed that the presmoothing yields efficient estimates with better mean squared error than the classical least-squares estimates in the case of the linear regression model. Our simulation experiments indicate that the SMD approach may share this feature, in nonlinear settings.

III An iterative SMD approach

The SMD method seems valuable compared to existing methods that apply to CEE models as in equation (I.1) and guarantee consistency. However, it is based on the optimization of a nonlinear contrast. To avoid a complicated optimization problem, here we present an iterative approach of SMD, based on a quadratic approximation of the nonlinear contrast. This iterative estimator is easy to compute and preserves all the benefits of the SMD estimator.

Let

$$\nabla_\beta g(Z,\beta) = \frac{\partial g}{\partial \beta}(Z,\beta)$$

be the $(p \times r)-$matrix of first order derivatives of g with respect to the components

of β. For β' close to β, we can write

$$g(Z,\beta) \approx g(Z,\beta') + \nabla_\beta g(Z,\beta')^\top(\beta-\beta') \stackrel{def}{=} q(Z,\beta,\beta'). \tag{III.1}$$

Thus, any β' close to β defines a linear approximation $\beta \mapsto q(Z,\beta,\beta')$ for $g(Z,\beta)$. Then it is natural to define the iterations

$$\beta^{(k)} = \arg\min_{\beta\in\mathcal{B}} M_{n,h}(\beta,\beta^{(k-1)}), \quad k=1,2,\ldots \tag{III.2}$$

where

$$M_{n,h}(\beta,\beta') = \frac{1}{2n^2}\sum_{1\le i,j\le n} g(Z_i,\beta,\beta')^\top g(Z_j,\beta,\beta')K_{ij},$$

$\beta^{(0)}$ is some initial value and $K(\cdot)$ has a positive Fourier Transform.

At each step k, the optimization problem (III.2) has an explicit solution:

$$\begin{aligned}\beta^{(k)} &= \beta^{(k-1)} - \left[\alpha_n I_p + \tfrac{1}{n^2}\textstyle\sum_{1\le i,j\le n}\nabla_\beta g(Z_i,\beta^{(k-1)})\nabla_\beta g(Z_j,\beta^{(k-1)})^\top K_{ij}\right]^{-1} \\ &\quad\times\left[\tfrac{1}{n^2}\textstyle\sum_{1\le i,j\le n}\nabla_\beta g(Z_i,\beta^{(k-1)})g(Z_j,\beta^{(k-1)})K_{ij}\right],\ k=1,2,\ldots,\end{aligned}$$

where I_p is a $p\times p-$identity matrix and $\alpha_n>0$ is a regularization parameter, decreasing to zero. The role of this parameter is to avoid the inversion of ill-conditioned matrices. In general, the positiveness of the Fourier Transform $\mathcal{F}[K]$ guarantees that the smallest eigenvalue of the $(p\times p)-$matrix

$$\mathbb{E}\left[\frac{1}{n^2}\sum_{1\le i,j\le n}\nabla_\beta g(Z_i,\beta)\nabla_\beta g(Z_j,\beta)^\top K_{ij}\right]$$

stays away from zero for $\beta\in\mathcal{B}$. However, our simulation experience indicates that a positive regularization parameter α_n may lead to more stable iterations. The estimator we propose is

$$\widehat{\beta} = \beta^{(k^*)},$$

for some value k^* obtained from a stopping rule for the iterations. We call it an *iterative SMD* estimator. Like $\tilde{\beta}_{n,h}$, the iterative estimator $\widehat{\beta}$ depends on h.

Let us provide an alternative point of view for the iterative SMD. Instead of

minimizing the criterion $M_{n,h}(\beta)$ to obtain the SMD estimator, one could instead search the solutions of the equations $\nabla_\beta M_{n,h}(\beta) = 0$, *i.e.* one could solve the system

$$F_n(\beta) = \frac{1}{n^2} \sum_{1 \leq i,j \leq n} \nabla_\beta g(Z_i, \beta) g(Z_j, \beta) K_{ij} = 0, \tag{III.3}$$

where $F_n : \mathcal{B} \subset \mathbb{R}^p \to \mathbb{R}^p$ is a vector-valued function of β that depends on the sample. Next, one could apply a numerical method to solve the system (III.3), like for instance the Newton-Kantorovich method, or one of its variants. The Newton-Kantorovich method consists in considering the iterations

$$\beta^{(k)} = \beta^{(k-1)} - \left[\nabla_\beta F_n(\beta^{(k-1)})\right]^{-1} F_n(\beta^{(k-1)}), \qquad k = 1, 2, \cdots$$

See Kantorovich & Akilov (1964). Moreover, one could consider the Levenberg-Marquardt methods to build the iterations

$$\beta^{(k)} = \beta^{(k-1)} - \left[\gamma_k I_p + \nabla_\beta F_n(\beta^{(k-1)})\right]^{-1} F_n(\beta^{(k-1)}), \qquad k = 1, 2, \cdots$$

See Levenberg (1944), Marquardt (1963), see also Goldfeld *et al.* (1966). The choice of γ_k is discussed, for instance, in Ortega & Rheinboldt (1970). One can easily remark that our procedure is closely related to a Newton-Kantorovich type method. In our case we use an approximation of the matrix $\nabla_\beta F_n(\beta)$, that is expected to be very accurate when β is close to β_0. Hence, in some sens, the iterative SMD is the limit of a Newton-Kantorovich type method starting from the first order conditions of the SMD approach. Thus one could expect that the SMD estimator and the iterative SMD estimator to be asymptotically equivalent.

The idea of the iterative SMD estimator could be extended to the case where $g(\cdot, \beta)$ is replaced by a weighted version build with some positive definite weight $(r \times r)-$matrix $W(\cdot, \beta)$, for instance to achieve asymptotic efficiency. It suffices to redefine the function $q(\cdot, \beta, \beta')$ from equation (III.1) as

$$q(Z, \beta, \beta') = W^{-1/2}(Z, \beta') \left\{ g(Z, \beta') + \nabla_\beta g(Z, \beta')^\top (\beta - \beta') \right\}$$

and redefine the optimization problem and the iterations accordingly. More precisely, consider

$$\beta^{(k)} = \beta^{(k-1)} - \Bigg[\alpha_n I_p + \frac{1}{n^2}\sum_{1\le i,j\le n} \nabla_\beta g(Z_i,\beta^{(k-1)}) W^{-1/2}(Z_i,\beta^{(k-1)}) W^{-1/2}(Z_j,\beta^{(k-1)}) \nabla_\beta g(Z_j,\beta^{(k-1)})^\top K_{ij}\Bigg]^{-1} \times \left[\frac{1}{n^2}\sum_{1\le i,j\le n} \nabla_\beta g(Z_i,\beta^{(k-1)}) W^{-1/2}(Z_i,\beta^{(k-1)}) W^{-1/2}(Z_j,\beta^{(k-1)}) g(Z_j,\beta^{(k-1)}) K_{ij}\right],$$

$k = 1, 2, \ldots$

Clearly, one could use the linearization idea to build an iterative version of the bootstrap estimator $\tilde{\beta}^*_{n,h}$ as solution of the quadratic optimization problem

$$\beta^{(k),*} = \arg\min_{\beta\in\mathcal{B}} M^*_{n,h}(\beta, \beta^{(k-1),*}), \quad k = 1, 2, \ldots, \tag{III.4}$$

where

$$M^*_{n,h}(\beta, \beta^{(k-1),*}) = \frac{1}{2n^2}\sum_{1\le i,j\le n} q(Z_i,\beta,\beta^{(k-1),*})^\top q(Z_j,\beta,\beta^{(k-1),*}) K^*_{ij},$$

with K^*_{ij} defined as in equation (II.4). The initial value $\beta^{(0),*}$ could be taken equal to $\widehat{\beta}$, and the stopping rule could be the same as for $\widehat{\beta}$, yielding the bootstrap iterative estimate $\widehat{\beta}^*$.

It is worthwhile to notice that, at each step, $\beta^{(k)}$ is obtained as the solution of a weighted least squares regression problem. Thus one can extend the scope of the SMD to a penalized SMD procedure inspired by the existing penalized least squares approaches. More precisely, one can modify the iterations in equation (III.2) to the following ones :

$$\beta^{(k)} = \arg\min_{\beta\in\mathcal{B}} \left\{M_{n,h}(\beta, \beta^{(k-1)}) + p_\lambda(\beta)\right\}, \quad k = 1, 2, \ldots, \tag{III.5}$$

where $p_\lambda(\cdot)$ is some penalty function depending on some tuning parameter $\lambda > 0$. Let us mention three types of penalties that are commonly used in the literature.

Perhaps the most known one is the ℓ_1-penalty function $p_\lambda(\beta) = \lambda|\beta|_1$, resulting in the least absolute shrinkage and selection operator (LASSO); see Tibshirani (1996). Next, let us recall the hard thresholding penalty function

$$p_\lambda(\beta) = \lambda^2 - (|\beta|_1 - \lambda)^2 \mathbf{1}\{|\beta|_1 < \lambda\};$$

see Antoniadis (1997) and Fan (1997). Here and in the following, $\mathbf{1}\{\cdot\}$ denotes the indicator function of the event in the curly brackets. More recently, Fan & Li (2001) investigated the smoothly clipped absolute deviation (SCAD) penalty function

$$p_\lambda(\beta) = \begin{cases} \lambda|\beta|_1 & \text{if} \quad 0 \leq |\beta|_1 < \lambda, \\ \frac{(\alpha^2-1)\lambda^2 - (|\beta|_1 - \alpha\lambda)^2}{2(\alpha-1)} & \text{if} \quad \lambda \leq |\beta|_1 < \alpha\lambda, \\ \frac{(\alpha+1)\lambda^2}{2} & \text{if} \quad |\beta|_1 \geq \alpha\lambda, \end{cases}$$

where α is an additional tuning parameter. On the basis on their empirical experience, Fan & Li (2001) propose to take $\alpha = 3.7$.

IV Simulation experiments

In this section we present empirical results from several simulation setup we considered in order to investigate the performance of our iterative approach. The first model we consider is the common logistic regression. We compare the results obtained by our method with those yield by the maximum likelihood method. The second example is linear regression model with an endogenous regressor. In this case we compare our method with the classical GMM. We end the section with a penalized regression example. In all our experiments we use a standard Gaussian kernel $K(\cdot)$.

IV.1 Logistic regression

In this experiment the data $Z_i = (Y_i, X_i^\top)^\top$, $1 \leq i \leq n$, with $Y_i \in \{0, 1\}$ and $X_i \in \mathbb{R}^3$, are independently generated from the model

$$Y = \mathbf{1}\{Y^\flat > 0\} \qquad \text{where} \qquad Y^\flat = X^\top \beta_0 + \varepsilon \tag{IV.6}$$

with ε independent of X and having a logistic law, *i.e.* $\mathbb{P}(\varepsilon \leq t) = \exp(t)/\{1 + \exp(t)\}$, $t \in \mathbb{R}$, and

$$X \sim N_3 \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{bmatrix} \right).$$

The true value of the parameter $\beta = (\beta_1, \beta_2, \beta_3)^\top$ is $\beta_0 = (1.2, 0.6, 0.8)^\top$. Here we consider

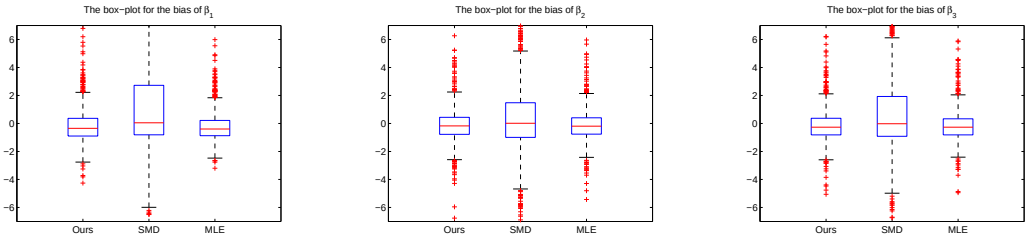
$$g(Z, \beta) = \left(Y - \frac{\exp(X^\top \beta)}{1 + \exp(X^\top \beta)} \right) X,$$

which corresponds to the score equations. In all experiments, the starting point $\beta^{(0)}$ for the iterative SMD was the vector $(1, 1, 1)^\top$. The regularization parameter α_n is set equal to 0.01 and the stopping rule for the iterations is $\|\beta^{(k)} - \beta^{(k-1)}\| / \|\beta^{(k-1)}\| < 0.0001$.

Here we compare the iterative SMD method with the maximum likelihood estimator (MLE) and SMD estimator. For sample sizes $n = 20$ and $n = 30$ we generate 1000 independent data sets. The simulation results are shown in the Table 2.1, while box-plots of the estimators are presented in the Figure 2.1 and 2.2. The the bandwidth used in simulation is selected in a grad $\{0.6, 0.2, \ldots, 1.5\}$ such that the value of $M_{n,h}(\widehat{\beta})$ is minimal. Overall, our iterative SMD performs well with small samples, compared to the MLE, that is known to be asymptotically efficient.

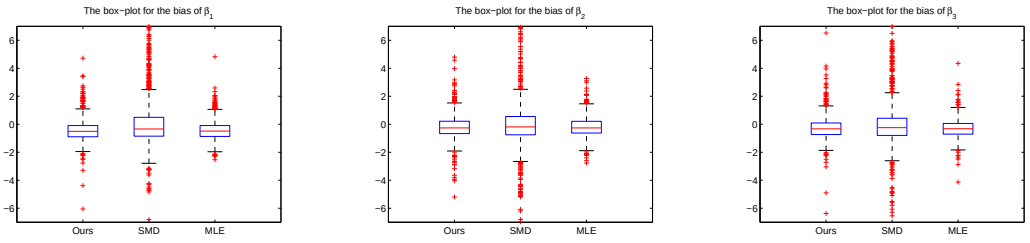
Table 2.1: The simulation results for the estimator of parameter obtained from 1000 replications generated using the model (IV.6)

			mean	std	mse
$n = 20$	$\beta_1 = 1.2$	Ours	1.0515	1.2074	1.4783
		MLE	1.0078	1.2186	1.5204
		SMD	3.2550	6.3594	44.6248
	$\beta_2 = 0.6$	Ours	0.4790	1.2835	1.6604
		MLE	0.4990	1.2578	1.5906
		SMD	1.7603	5.7624	34.5189
	$\beta_3 = 0.8$	Ours	0.6687	1.2637	1.6126
		MLE	0.6348	1.1253	1.2924
		SMD	2.1462	5.9365	37.0195
$n = 30$	$\beta_1 = 1.2$	Ours	0.7879	0.8463	0.8853
		MLE	0.7713	0.6926	0.6629
		SMD	1.7388	3.7282	14.1759
	$\beta_2 = 0.6$	Ours	0.3996	0.8737	0.8028
		MLE	0.4002	0.7236	0.5630
		SMD	1.0146	3.6281	13.3218
	$\beta_3 = 0.8$	Ours	0.5263	0.8174	0.7424
		MLE	0.5008	0.6626	0.5281
		SMD	1.2067	3.4738	12.2205



(a) (b) (c)

Figure 2.1: Box-plots of estimated coefficients in the model (IV.6), sample size $n = 20$: the left box-plots are for the iterative SMD method, the middle one for SMD and the right ones for the MLE.



(a) (b) (c)

Figure 2.2: The same results as in Figure 2.1 for the sample size $n = 30$

IV.2 Endogeneity

The SMD approach offers an appealing way to estimate nonlinear regression models with endogenous regressors. Consider a nonlinear regression model

$$Y = m(W, \beta) + u, \qquad \text{but} \qquad \mathbb{E}[u \mid W] \neq 0 \ \ a.s.,$$

where $Y \in \mathbb{R}$, $W \in \mathbb{R}^d$. Meanwhile, assume that one observes a vector $X \in \mathbb{R}^q$ such that $\mathbb{E}[u \mid X] = 0$. The vector X is the so-called vector of instruments. Then, one could rewrite these conditions under the form

$$\mathbb{E}[g(Z, \beta) \mid X] = 0 \ a.s. \quad \text{with} \quad Z = (Y, W^\top, X^\top)^\top \quad \text{and} \quad g(Z, \beta) = Y - m(W, \beta),$$

and apply the SMD approach.

In this section we consider two such situations in the case of the regression model

$$Y = (W^\top \beta)^2 + u, \tag{IV.7}$$

with endogenous vector of $W = (W_1, W_2)^\top \in \mathbb{R}^2$ and the true value of the parameter $\beta_0 = (0.5, 0.9)^\top$. In the first setup, we consider a variable U and a real-valued X such that

$$\begin{bmatrix} u \\ U \\ X \end{bmatrix} \sim N_3 \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right),$$

and set

$$W_1 = X + U/2 \quad \text{and} \quad W_2 = X^2 + U.$$

In the second setup we define the endogenous regressors $W = (W_1, W_2)^\top$ as $W_1 =$

$X_1 + X_2 + V$ and $W_2 = X_1 X_2 + V$, where V and $X = (X_1, X_2)^\top$ are defined as

$$\begin{bmatrix} u \\ V \\ X_1 \\ X_2 \end{bmatrix} \sim N_4 \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right).$$

For both setups, we repeat our simulation experiment 1000 times with samples of size $n = 20$ and $n = 30$. We compare our method with the standard GMM and SMD. For the first setup we use the optimal instruments that one could derive knowing the structure of the endogenous regressors and the true value of the parameters. (See Newey (1993) for the general form of the optimal instruments.) Such instruments are in general infeasible in applications where one does not know the endogeneity structure, and thus represent a benchmark. Thus, in the first setup we apply GMM with the estimating equations

$$\begin{cases} \mathbb{E}\left[\{1.8X^3 + X^2 + 1.15\}\{Y - (W^\top \beta)^2\}\right] = 0 \\ \mathbb{E}\left[\{1.8X^4 + X^3 + 2.3\}\{Y - (W^\top \beta)^2\}\right] = 0 \end{cases}.$$

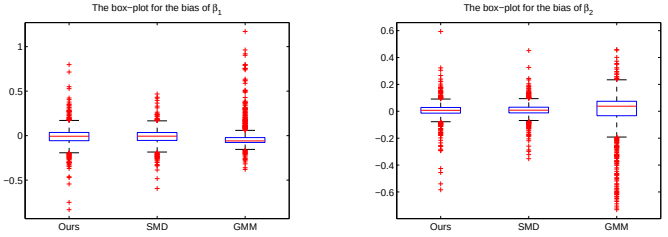
In the second setup, the optimal instruments yield the estimating equations

$$\begin{cases} \mathbb{E}\left[\{(X_1 + X_2)^2 + 1.8X_1X_2(X_1 + X_2) + 1.15\}\{Y - (W^\top \beta)^2\}\right] = 0 \\ \mathbb{E}\left[\{X_1X_2(X_1 + X_2) + 1.8X_1^2X_2^2 + 2.3\}\{Y - (W^\top \beta)^2\}\right] = 0. \end{cases},$$

We present the mean, the standard deviation and the mean squared error of the of the three types of estimators in the Table 2.2 and 2.3. Figure 2.3 to 2.6 are the box-plot figures. We use a data-driven bandwidth rule by minimization of $M_{n,h}(\widehat{\beta})$ on the grid $\{0.1, 0.2, \ldots, 0.5\}$, as described in the logistic regression example. We can see that the results of our iterative SMD approach are better than the classical GMM approach, and close to SMD.

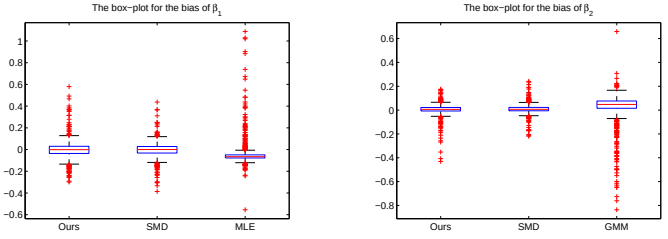
Table 2.2: The simulation results for model (IV.7)with the first setup.

			mean	std	mse
$n = 20$	$\beta_1 = 0.5$	Ours	0.4892	0.1126	0.0128
		GMM	0.4787	0.1410	0.0203
		SMD	0.4881	0.0919	0.0086
	$\beta_2 = 0.9$	Ours	0.9063	0.0697	0.0049
		GMM	0.8947	0.1584	0.0251
		SMD	0.9097	0.0563	0.0033
$n = 30$	$\beta_1 = 0.5$	Ours	0.4977	0.0787	0.0062
		MLE	0.4556	0.1118	0.0144
		SMD	0.4958	0.0665	0.0044
	$\beta_2 = 0.9$	Ours	0.9054	0.0454	0.0021
		MLE	0.9247	0.1170	0.0143
		SMD	0.9089	0.0372	0.0015



(a) (b)

Figure 2.3: Box-plots of estimated coefficients in the model (IV.7) with the first setup, sample size $n = 20$: the left box-plots are for the iterative SMD method, the middle one for SMD and the right ones for the GMM.

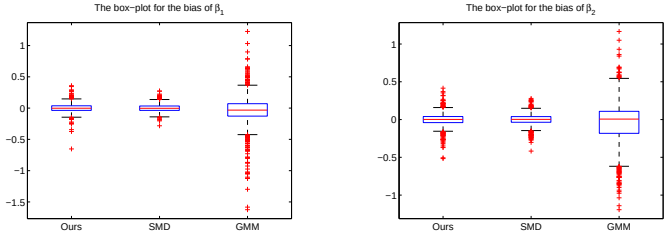


(a) (b)

Figure 2.4: The same results as in Figure 2.3 for the sample size $n = 30$

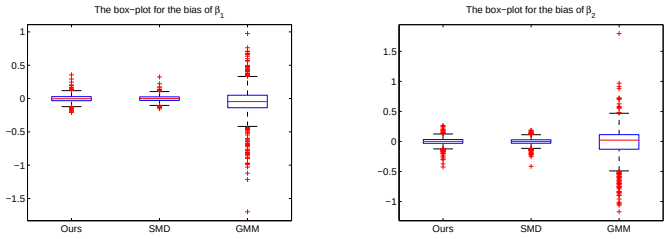
Table 2.3: The simulation results for model (IV.7)with the second setup.

			mean	std	mse
$n = 20$	$\beta_1 = 0.5$	Ours	0.5006	0.0726	0.0053
		GMM	0.4540	0.2719	0.0760
		SMD	0.5004	0.0611	0.0037
	$\beta_2 = 0.9$	Ours	0.8986	0.0881	0.0078
		GMM	0.48455	0.3085	0.0981
		SMD	0.8989	0.0759	0.0058
$n = 30$	$\beta_1 = 0.5$	Ours	0.4991	0.0543	0.0029
		GMM	0.4392	0.2466	0.0644
		SMD	0.5003	0.0453	0.0021
	$\beta_2 = 0.9$	Ours	0.8990	0.0640	0.0041
		GMM	0.8655	0.2869	0.0834
		SMD	0.8976	0.0548	0.0030



(a) (b)

Figure 2.5: Box-plots of estimated coefficients in the model (IV.7) with the second setup, sample size $n = 20$: the left box-plots are for the iterative SMD method, the middle one for SMD and the right ones for the GMM.



(a) (b)

Figure 2.6: The same results as in Figure 2.5 for the sample size $n = 30$

IV.3 Penalized regression

In this example, we simulate 200 independent samples of 40 i.i.d. observations from the model

$$Y = (X^\top \beta_0)^2 + \sigma(X)\varepsilon \tag{IV.8}$$

where $\beta_0 = (3, 1.5, 0, 0, 2, 0, 0, 0)^\top$, $\sigma^2(X)$ is the conditional variance of Y and ε is standard normal independent of X. The components of $X = (X_1, \ldots, X_8)^\top$ are standard normal but correlated: the correlation between X_i and X_j is $\rho^{|i-j|}$ with $\rho = 0.5$. We consider several choices for $\sigma(X)$ corresponding to homoscedastic and heteroscedastic error terms: $\sigma(X) \equiv 1, 3, 9$, and $\sigma(X) = (X^\top \beta_0)^2/5$.

Here we use the ℓ_1-penalty function $p_\lambda(\beta) = \lambda|\beta|_1$ in our penalized criterion (III.5), where $\lambda \geq 0$ is a penalty parameter. We select λ by minimization of

$$\frac{1}{n^2} \sum_{1 \leq i,j \leq n} g(Z_i, \widehat{\beta}) g(Z_j, \widehat{\beta}) K_{ij}$$

on a grid of points $\{0.5, 1, \cdots, 5\}$.

In the Table 2.4, we show the mean and the standard deviation of the iterative SMD estimator with a penalized criterion and the LASSO. 'Correct' indicates the average number of estimates of β_3, β_4, β_6, β_7, β_8 that are equal to zero in each of the 200 experiments. The ideal number is 5. 'Incorrect' indicates the average of coefficients β_1, β_2, β_5, erroneously set to 0 by our estimation procedure.

Table 2.4: The iterative SMD with a ℓ_1-penalty function applied to the model (IV.8). The results are the averages obtained with 200 samples of size $n = 40$.

$\sigma(X)$		1	3	9	$(X^\top\beta_0)^2/5$
$\beta_1 = 3$	Ours	2.996(0.019)	2.996(0.042)	2.993(0.125)	2.984(0.314)
	LASSO	2.996(0.020)	2.997(0.043)	2.991(0.118)	2.933(0.166)
$\beta_2 = 1.5$	Ours	1.490(0.021)	1.481(0.050)	1.448(0.139)	1.354(0.363)
	LASSO	1.497(0.023)	1.482(0.049)	1.455(0.139)	1.433(0.151)
$\beta_3 = 0$	Ours	0.000(0.004)	0.005(0.022)	0.032(0.069)	0.076(0.160)
	LASSO	0.002(0.011)	0.014(0.030)	0.049(0.086)	0.036(0.069)
$\beta_4 = 0$	Ours	0.000(0.005)	0.004(0.016)	0.028(0.056)	0.077(0.158)
	LASSO	0.001(0.008)	0.010(0.025)	0.047(0.092)	0.026(0.058)
$\beta_5 = 2$	Ours	1.981(0.017)	1.966(0.041)	1.917(0.123)	1.796(0.316)
	LASSO	1.992(0.020)	1.977(0.043)	1.939(0.128)	1.901(0.127)
$\beta_6 = 0$	Ours	0.000(0.005)	0.006(0.022)	0.040(0.073)	0.085(0.177)
	LASSO	0.003(0.015)	0.012(0.028)	0.050(0.078)	0.036(0.069)
$\beta_7 = 0$	Ours	0(0)	0.002(0.013)	0.017(0.048)	0.038(0.097)
	LASSO	0.001(0.008)	0.005(0.020)	0.031(0.063)	0.027(0.065)
$\beta_8 = 0$	Ours	0(0)	0.002(0.011)	0.022(0.046)	0.057(0.127)
	LASSO	0.001(0.009)	0.006(0.020)	0.044(0.068)	0.021(0.052)
Correct	Ours	4.975	4.695	3.805	3.67
	LASSO	4.68	4.3	3.37	3.755
Incorrect	Ours	0	0	0	0
	LASSO	0	0	0	0

V References

1. Ai, C., and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* **71**, 1795-1843.

2. Antoine, B., Bonnal, H., and Renault, E. (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood. *J. Econometrics* **138**, 461-487.

3. Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion). *J Ital. Statist. Assoc.* **6**, 97-130.

4. Carrasco, M., and Florens, J.P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory* **16**, 797-834.

5. Cristóbal-Cristóbal, J. A., Faraldo-Roca, P. and González-Manteiga, W. (1987). A class of linear regression parameter estimators constructed by nonparametric estimation. *Ann. Statist.* **15**, 603-609.

6. Dominguez, M.A., and Lobato, I.N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica* **72**, 1601-1615.

7. Donald, S.G., Imbens, G.W., and Newey, W.K. (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *J. Econometrics* **117**, 55-93.

8. Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. **96**, 1348-1360.

9. Fan, J. (1997). Comment on "Wavelets in statistics: a review" (with discussion). *J Ital. Statist. Assoc.* **6**, 97–144.

10. Faraldo-Roca, P. and González-Manteiga, W. (1987). On efficiency of a new class of linear regression estimates obtained by preliminary nonparametric estimation. In *New Perspectives in Theoretical and Applied Statistics* (Edited by M.L. Puri, J. Perez-Vilaplana and W. Wertz), 229-242. Wiley,

New York.

11. Goldfeld, S., Quandt, R., and Trotter H. (1966). Maximization by quadratic hill climbing. *Econometrica* **34**, 541-551.

12. Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometric* **50**, 1029-1054.

13. Hjort, N.L., Mckeague, I.W., and Van Keilegon, I. (2009). Extending the sscope of empirical likelihood. *Ann. Statist.* **37**, 1079-1111.

14. Kantorovich, L.V., and Akilov, G.P. (1964), *Functional Analysis in Normed Spaces*. Pergamon Press.

15. Kitamura, Y., Tripathi, G., and Ahn, H. (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* **72**, 1667-1714.

16. Lavergne, P., and Patilea, V. (2013). Smooth minimum distance estimation and testing in conditional moment restrictions models: uniform in bandwidth theory. *J. Econometrics* **177**, 47-59.

17. Levenberg, K. (1944). A method for the solution of certain nonlinear problems in least squares. *Quart. Appl. Math.* **2**, 164-168.

18. Marquardt, D. (1963). An algorithm for least squares estimation of nonlinear parameters. *SIAM J. Appl. Math.* **11**, 431-441.

19. Newey W.K. (1993). Efficient Estimation of Models with Conditional Moment Restrictions. in *Handbook of Statistics*, Vol. 11, ed. by G.S. Maddala, C.R. Rao, and H.D. Vinod. Amsterdam: North-Holland, Chapter 16.

20. Ortega, J.M., and Rheinboldt, W.C. (1970). *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York London.

21. Qin, J., and Lawless, J. (2004). Empirical likelihood and generalized estimating equations. *Ann. Statist.* **22**, 300-325.

22. Smith, R.J. (2007a). Efficient information theoretic inference for conditional moment restrictions. *J. Econometrics* **138**, 430-460.

23. Smith, R.J. (2007b). Local GEL estimation with conditional moment restrictions. In *The refinement of econometric estimation and test procedures: finite sample and asymptotic analysis*, eds. G.D.A. Phillips and E. Tzavalis, 100-122. Cambridge Univeersity Press: Cambridge.

24. Tibshirani, R.J. (1996). Regression shrinkage and selection via the LASSO. *J. Royal. Statist. Soc B* **58**, 267-288.

3

A new inference approach for single-index models

I Introduction

Modeling the relationship between one or several response variables and a vector of covariates is a common problem in statistics. Usually, one aims to model the conditional law of the responses given the covariates, or at least some characteristics of this conditional law, such as the mean, the median, the higher order moments, *etc.* In a parametric approach, one specifies a set indexed by a vector of parameters, *i.e.* a model, to which is supposed to belong the conditional law, or its characteristic of interest. The linear regression is the most prominent example. In a fully nonparametric approach, the model is specified as large as possible, but of course there is a price for the model complexity reflected in the poor accuracy of the estimators. Therefore, often one looks for semiparametric approaches that realize a compromise between the accuracy that could be obtained in parametric model and the flexibility of nonparametric specifications.

Single-index models are common semiparametric approaches that realize such a compromise. The underlying assumption is that the information on the quantity of interest (the law of the responses or some characteristics of it) carried by the vector of covariates is the same as the information carried by a one-dimensional projection of the covariate vector, the so-called *index*. In other words, one still consider a nonparametric approach, but only after a dimension reduction step which replaces the original covariate vector by some linear combination of its components. See, for instance, Powell *et al.* (1989), Klein & Spady (1993), Ichimura (1993), Härdle *et al.*

(1993), Carroll *et al.* (1997), Hristache *et al.* (2001), Yin & Cook (2002), Kong & Xia (2007), Horowitz (2009), Liang *et al.* (2010), Kong & Xia (2012), Ma & Zhu (2013), and the references therein.

Despite the extensive literature on single-index models, some technical aspects remain unsatisfactorily solved. For instance, in most contributions the covariates are supposed to have a bounded support. Even with bounded support covariates, a trimming is usually employed to keep density estimates, usually appearing in the denominators, away from zero. In many contributions considering the additive regression setup, the error term is supposed homoscedastic.

In this chapter we introduce a new general inference approach for the index in conditional models using a single-index assumption. Our approach is based on kernel smoothing and could be applied to any existing framework, under a mild condition. For simplicity, here we focus on the single-index mean regressions and single-index conditional law cases. We allow for unbounded discrete and continues covariates, for heteroscedastic error terms in the mean regression setup, and no trimming is involved in the inference. The approach follows and extends the idea of the Smooth Minimum Distance estimation method of Lavergne & Patilea (2013) from parametric to a semiparametric framework.

The paper is organized as follows. The underlying idea of the new approach is presented in section II. The corresponding estimators are introduced in section III and their consistency and asymptotic normality are derived. In section IV we propose a simple procedure for constructing confidence intervals for the index coefficients. Some empirical evidence on the performance of our inference method is provided in section V, using both simulated and real data examples. The simulation results indicate that our method performs well compared to existing approaches. The technical aspects are postponed to the Appendix.

II The framework

Assume that the observations are independent copies of $(Y^\top, X^\top)^\top$ where $Y \in \mathbb{R}^d$, $d \geq 1$, denote the random response vector and $X \in \mathbb{R}^p$, $p \geq 1$, stands for the

random column vector of covariates. (Here and in the following, for a matrix A, $A^\top$ stands for the transpose of A.) For mean regression the single-index assumption means that there exists a column parameter vector $\beta_0 \in \mathbb{R}^p$ such that

$$\mathbb{E}[Y \mid X] = \mathbb{E}[Y \mid X^\top \beta_0]. \tag{II.1}$$

The scalar product $X^\top \beta_0$ is the so-called index. The direction β_0 and the nonparametric univariate regression $\mathbb{E}[Y \mid X^\top \beta_0]$ have to be estimated. See Hristache *et al.* (2001), Delecroix *et al.* (2006), Horowitz (2009), Cui *et al.* (2011) and the references therein for a panorama of the existing estimation procedures. When applying the single-index paradigm to conditional laws of Y given X, one supposes

$$Y \perp X \mid X^\top \beta_0. \tag{II.2}$$

In this case the direction defined by β_0 and the conditional law of the response Y given the index $X^\top \beta_0$ have to be estimated. See Delecroix *et al.* (2003), Hall & Yao (2005), Chiang & Huang (2012), Zhang *et al.* (2015) for the available estimation approaches. In both situations only the direction given by β_0 is identified, so that a suitable identification condition should accompany the model assumption.

In order to formulate the problem in an unified way, consider T_u, $u \in \mathcal{U}$, a family of transformations of the response variable Y. The transformations T_u take values in some finite dimensional space that could be, for instance, the space of Y or the real line. The set $\mathcal{U}$ is contained in some finite dimensional space. Then, the general single-index model (SIM) assumption we consider is the following:

$$\text{there exists a unique } \beta_0 \text{ such that} \quad E[T_u(Y)|X] = E[T_u(Y)|X^\top \beta_0], \quad \forall u \in \mathcal{U}, \tag{II.3}$$

where β_0 is an unknown index vector which belongs to the parameter set

$$\mathcal{B} \subset \{(\beta_1, \dots \beta_p) : \beta_1 = 1\} \subset \mathbb{R}^p. \tag{II.4}$$

This framework allows to take into account the two single-index assumptions pre-

sented above. Indeed, if the family of transformation contains only the identity transformation, *i.e.,* $T_u(y) = y$ for any $u \in \mathcal{U}$, one recovers the condition (II.1). Meanwhile, if $T_u(y) = \mathbf{1}\{y \leq u\}$, $u \in \mathcal{U}$, with $\mathcal{U} = \mathbb{R}^p$, then (II.3) becomes equivalent to condition (II.2). (Here and in the following, for any v_1 and v_2 vectors of the same dimension, $v_1 \leq v_2$ stands for the vector componentwise inequality between v_1 and v_2.) For simplicity, hereafter we only consider the condition (II.3) for one of these two type of transformations T_u.

Let us assume that for any $\beta \in \mathcal{B}$, the random variable $X^\top \beta$ has a density denoted by $f_\beta(\cdot)$. To estimate a parameter β_0 that satisfies the condition (II.3), first let us define

$$g_u(Y, z; \beta) = \left\{ T_u(Y) - E[T_u(Y) | X^\top \beta = z] \right\} f_\beta(z), \qquad z \in \mathbb{R},\ \beta \in \mathcal{B},\ u \in \mathcal{U}. \tag{II.5}$$

Then, condition (II.3) is equivalent to the following one:

$$\mathbb{E}[g_u(Y, X^\top \beta; \beta) \mid X] = 0 \quad \text{almost surely}, \quad \forall u \in \mathcal{U} \quad \Leftrightarrow \quad \beta = \beta_0. \tag{II.6}$$

Next, the idea is to build a contrast function that allows to encompass the conditional moment conditions in one marginal (unconditional) equation. For this purpose, let $(Y_1^\top, X_1^\top)^\top$ and $(Y_2^\top, X_2^\top)^\top$ be two independent copies of $(Y^\top, X^\top)^\top$ and let $\omega(\cdot)$ be a real-valued integrable function defined on the space of X. Assume that $\omega(\cdot)$ has an integrable, strictly positive Fourier Transform. For instance, one could take $\omega(x) = \exp(-\|x\|^2/2)$, $x \in \mathbb{R}^p$. Finally, define the real-valued contrast function

$$Q(\beta) = \int_{\mathcal{U}} \mathbb{E}[g_u(Y_1, X_1^\top \beta; \beta)^\top g_u(Y_2, X_2^\top \beta; \beta) \omega(X_1 - X_2)] d\mu(u), \quad \beta \in \mathcal{B}, \tag{II.7}$$

where μ is some probability measure with support $\mathcal{U}$ considered with the Borel $\sigma-$field. As it will be mentioned in the following, for the case corresponding to the assumption (II.2), a convenient choice is $\mu = F_Y$, where F_Y denotes the probability distribution of Y. In applications F_Y is unknown and could be replaced by a given approximation or by the empirical distribution of the sample of Y.

The following result guarantees the direction β_0 from the condition (II.6) could

be identified as the unique root of the contrast $Q(\cdot)$.

Lemma II.1. *Let $\mathcal{B}$ be some parameter set defined as in equation (II.4). Assume that the Fourier Transform of $\omega(\cdot)$ is strictly positive and integrable. Then $Q(\beta) \geq 0$, $\forall \beta \in \mathcal{B}$. Moreover, condition (II.6) holds true if and only if $Q(\beta_0) = 0$ and $Q(\beta) > 0$ for any $\beta \neq \beta_0$.*

The idea of our estimation approach is to build a sample based approximation of $Q(\beta)$ and to minimize it with respect to the parameter β. Let us point out that, by the definition of the functions g_u, the covariates will be allowed to have unbounded support and no trimming will be necessary in the approximation of $Q(\beta)$.

Let us point out that, in general, one could not simply use a least-squares type contrast instead of $Q(\beta)$. For illustration, let us consider that case of a single-index assumption for the mean regression of a real-valued response, *i.e.,* $Y = \mathbb{E}[Y \mid X^\top \beta_0] + \varepsilon$ and $\mathbb{E}[\varepsilon \mid X] = 0$. Then one can decompose

$$\mathbb{E}\left[g_u^2(Y, X^\top\beta; \beta)\right] = \mathbb{E}\left[\left\{\mathbb{E}[Y \mid X^\top\beta_0] - \mathbb{E}[Y \mid X^\top\beta]\right\}^2 f_\beta^2(X^\top\beta)\right] + \mathbb{E}\left[\varepsilon^2 f_\beta^2(X^\top\beta)\right],$$

and thus it becomes clear that β_0 may not be the minimum of $\mathbb{E}\left[g_u^2(Y, X^\top\beta; \beta)\right]$. Our contrast $Q(\beta)$, inspired by the Smooth Minimum Distance estimation method introduced by Lavergne & Patilea (2013), avoids the identification problem for β_0, provided the condition (II.6) holds true.

Finally, let us notice that the definition of the criterion $Q(\beta)$, and hence the estimation approach that will be described in the following, could be extended to the case of a multiple index assumption. It suffices to replace the index $X^\top\beta$ by a multiple index $X^\top B$ where B is a $p \times q-$matrix, $1 \leq q < p$, and to reconsider the construction above. For simplicity, herein, we focus on single-index assumptions.

III The estimation method

Let $(Y_i^\top, X_i^\top)^\top$, $1 \leq i \leq n$, be an independent sample from $(Y^\top, X^\top)^\top$. Our estimator of β_0 is

$$\widehat{\beta} = \arg\min_{\beta \in \mathcal{B}} \widehat{Q}_n(\beta),$$

where

$$\widehat{Q}_n(\beta) = \int_{\mathcal{U}} \left[\frac{1}{n^2} \sum_{i,j=1}^{n} \widehat{g}_u(Y_i, X_i^\top \beta; \beta)^\top \widehat{g}_u(Y_j, X_j^\top \beta; \beta) \omega_{ij} \right] d\mu_n(u), \qquad \beta \in \mathcal{B},$$

$\widehat{g}_u$ is an estimate of g_u, $\omega_{ij} = \omega(X_i - X_j)$ and μ_n in some probability measure that may depend on n. For estimating g_u we use kernel smoothing and define

$$\begin{aligned} \widehat{g}_u(y, z; \beta) &= T_u(y)\widehat{f}_\beta(z) - \mathbb{E}[T_u(Y)\widehat{\ \mid X^\top \beta = z}]\widehat{f}_\beta(z) \\ &= \frac{1}{nh} \sum_{k=1} \{T_u(y) - T_u(Y_k)\} K((X_k^\top \beta - z)/h), \end{aligned}$$

where $K(\cdot)$ is an univariate kernel and h is the bandwidth. The choice of μ_n matters only in the case of the single-index in law assumption (II.2), in which case we propose to consider the empirical distribution of the responses. More precisely, under the assumption (II.2) we propose

$$\widehat{Q}_n(\beta) = \frac{1}{n} \sum_{l=1}^{n} \frac{1}{n^2} \sum_{i,j=1}^{n} \widehat{g}_{Y_l}(Y_i, X_i^\top \beta; \beta) \widehat{g}_{Y_l}(Y_j, X_j^\top \beta; \beta) \omega_{ij}, \qquad \beta \in \mathcal{B}, \tag{III.8}$$

where

$$\widehat{g}_{Y_l}(Y_i, X_i^\top \beta; \beta) = \frac{1}{nh} \sum_{k=1}^{n} \{\mathbf{1}\{Y_i \le Y_l\} - \mathbf{1}\{Y_k \le Y_l\}\} K((X_k - X_i)^\top \beta / h).$$

In the case of the assumption (II.1) we propose the criterion

$$\widehat{Q}_n(\beta) = \frac{1}{n^2} \sum_{i,j=1}^{n} \widehat{g}(Y_i, X_i^\top \beta; \beta)^\top \widehat{g}(Y_j, X_j^\top \beta; \beta) \omega_{ij}, \qquad \beta \in \mathcal{B}, \tag{III.9}$$

where

$$\widehat{g}(Y_i, X_i^\top \beta; \beta) = \frac{1}{nh} \sum_{k=1}^{n} \{Y_i - Y_k\} K((X_k - X_i)^\top \beta / h).$$

Let us comment on a common feature of the single-index estimation methods. By the nature of the model, a nonparametric estimation is involved in any semiparametric single-index estimation approach. In general, this requires a control of small values of the nonparametric density estimators appearing in the denominators. A

common practice is to suppose that the density of $X^\top\beta$ is uniformly bounded away from zero for all $\beta \in \mathcal{B}$. Such a condition is quite unrealistic, even when X has a bounded support and a density bounded away from zero. Indeed, one may easily build a counterexample considering a bidimensional $X = (X_{(1)}, X_{(2)})^\top$ with two independent uniform random variables on $[0,1]$ and $\mathcal{B} = \{(1,\beta_2)^\top : |\beta_2| \leq b\}$, for some arbitrary $b > 0$. Then, except for $\beta_2 = 0$, the random variable $X_{(1)} + \beta_2 X_{(2)}$ does not have a density bounded away from zero. The usual remedy is to trim the criterion used for estimation, that is to remove the observations leading to small estimated values for the density of $X^\top\beta$. The trimming may be relaxed with the sample size, that is the fraction of removed observations could grow slower than the sample size, but one still has to use complicated arguments for the asymptotics. For both single-index situations we consider here, in mean and in law, the new approach we propose allows for unbounded covariates and does not require a trimming. To our best knowledge, our estimation method is the first one with this feature.

Let ∇_β the differential operator given by the last $(p-1)$ first order partial derivatives corresponding to the last $(p-1)$ components of β. In the case of the condition (II.2), let

$$J(\boldsymbol{\beta}_0) = \int_{\mathcal{U}} \mathbb{E}\Big\{\mathbb{E}[\nabla_\beta g_u(Y_1, X_1^\top\beta_0;\beta_0)\,|\,X_1]\mathbb{E}[\nabla_\beta g_u(Y_2, X_2^\top\beta_0;\beta_0)\,|\,X_2]^\top\omega(X_1 - X_2)\Big\}dF_Y(u),$$

$$\Sigma(\boldsymbol{\beta}_0) = 4\mathbb{E}\left[\psi(Y,X;\beta_0)\psi(Y,X;\beta_0)^\top\right],$$

and

$$\psi(Y_1,X_1;\beta_0) = \int_{\mathcal{U}} \mathbb{E}\Big\{\mathbb{E}\Big[\nabla_\beta g_u(Y, X^\top\beta_0;\beta_0)\,|\,X\Big]\,\omega(X - X_1)\,|\,X_1\Big\}g_u(Y_1, X_1^\top\beta_0;\beta_0)dF_Y(u).$$

In the case of single-index mean regression, $g_u(y,t;\beta)$ does not depend on u, hence the integrals with respect to F_Y, the probability distribution of the response, disappear from the definitions of the $(p-1)\times(p-1)-$matrices $J(\beta_0)$ and $\psi(Y,X;\beta_0)$ above. The following result describe the asymptotic behavior of the semiparametric estimator $\widehat{\beta}$. Below, $\rightsquigarrow$ denotes convergence in law and $\mathbf{0}_{p-1}$ is the null column vector in $\mathbb{R}^{p-1}$.

Proposition 3.1. *Let* $\widehat{\beta} = \arg\min_{\beta\in\mathcal{B}} \widehat{Q}_n(\beta)$ *for* $\widehat{Q}_n(\beta)$ *defined as in equation (III.8) or (III.9). Suppose that the identification condition (II.6) holds true. Under the Assumption VII.1,* $\widehat{\beta} \to \beta_0$, *in probability. If in addition Assumption VII.2 holds true,*

$$\sqrt{n}\left(\widehat{\beta} - \beta_0\right) \rightsquigarrow N_p(0, V),$$

where

$$V = \begin{pmatrix} 0 & \mathbf{0}'_{p-1} \\ \mathbf{0}_{p-1} & V_{p-1} \end{pmatrix} \qquad \textit{with} \qquad V_{p-1} = J(\boldsymbol{\beta}_0)^{-1}\Sigma(\boldsymbol{\beta}_0)J(\boldsymbol{\beta}_0)^{-1},$$

Given that $\widehat{\beta}$ is $\sqrt{n}-$consistent, one could derive the $\sqrt{nh}-$consistency for the conditional mean or the conditional distribution function of Y given X. This type of results are quite standard and straightforward, see for instance section 2.4 in Horowitz (2009), and hence will be omitted.

IV Confidence intervals

The asymptotic variance of $\widehat{\beta}$ has a complicated form. To approximate the law of $\widehat{\beta}$ with small and moderate samples, we propose a simulation based approach similar to the one proposed by Lavergne & Patilea (2013). See also Jin *et al.* (2001). The idea is to build a suitable randomly perturbed version of the criterion $\widehat{Q}_n(\beta)$ and to compute its minimum. Conditionally on the original sample, the law of this minimum is shown to be close to the law of $\widehat{\beta}$. Then it suffices to repeat the random perturbation procedure many times to derive a simulation based approximation of the law of $\widehat{\beta}$. More precisely, the steps of the procedure go as follows.

1. Generate a random sample $\xi_1, \cdots, \xi_n$ from a distribution with unit mean and unit variance, for instance the exponential law of parameter 1.

2. Build the randomly perturbed criterion

$$\widehat{Q}^*_n(\beta) = \int_{\mathcal{U}} \left[\frac{1}{n^2} \sum_{i,j=1}^{n} \widehat{g}_u(Y_i, X_i^\top\beta; \beta)^\top \widehat{g}_u(Y_j, X_j^\top\beta; \beta)\omega^*_{ij} \right] d\mu_n(u), \qquad \beta \in \mathcal{B},$$

where μ_n is the empirical distribution of the responses and $\omega_{ij}^* = \xi_i \xi_j \omega_{ij}$.

3. Define
$$\widehat{\beta}^* = \arg\min_{\beta \in \mathcal{B}} \widehat{Q}_n^*(\beta).$$

4. Repeat the above steps many times and approximate the law of $\widehat{\beta}$ using the sample of $\widehat{\beta}^*$'s.

The following result provides the asymptotic validity of this procedure. The arguments for the proof could be obtained by standard modifications of those for the proof of Proposition 3.1, and hence will be omitted.

Proposition 4.1. *Under the conditions of Proposition 3.1 guaranteeing the asymptotic normality of* $\sqrt{n}(\widehat{\beta} - \beta_0)$*, for any* $w \in \{0\} \times \mathbb{R}^{p-1}$,

$$\mathbb{P}\left(\sqrt{n}(\widehat{\beta}^* - \widehat{\beta}) \leq w \,|\, Y_1, X_1, \cdots, Y_n, X_n\right) - \mathbb{P}\left(\sqrt{n}(\widehat{\beta} - \beta_0) \leq w\right) \to 0, \quad \textit{in probability.}$$

V Empirical illustrations

We investigated the performance of our new approach to build parameter estimates and confidence intervals for single-index models through extensive simulation experiments and real data examples. The general conclusion is that our approach performs well, sometimes much better, compared with the existing approaches. In all our empirical studies we used a gaussian kernel $K(\cdot)$.

V.1 Simulation experiments with single-index in mean models

First, we consider two setups similar to the ones considered in Cui *et al.* (2011): the model equation is

$$Y = (X^\top \beta_0)^2 + \varepsilon, \tag{V.10}$$

with a three-dimensional vector of covariates $X = (X_{(1)}, X_{(2)}, X_{(3)})^\top$, where the independent sample of $(X_{(1)}, X_{(2)})^\top$ is generated from a bivariate normal law with mean equal to 1, standard deviations equal to 1 and correlation equal to 0.2. Meanwhile,

$X_{(3)}$ is a Bernoulli random variable with parameter $p = 0.4$. The true parameter is $\beta_0 = (\beta_{0,1}, \beta_{0,2}, \beta_{0,3})^\top = (1, 0.8, 0.5)^\top$. The first setup is a homoscedastic case where the error ε has a $N(0, 0.5^2)$ law. In this case the signal-to-noise ratio, that is SSR/SSE, is approximately equal to 76.6. In the second setup we introduce some heteroscedasticity by considering $\varepsilon \sim ((X^\top\beta)^2/5) * N(0, 1)$. Then the value of the signal-to-noise ratio is approximately equal to 13.

Our estimator $\widehat{\beta}$ depends on the bandwidth h. Here we select h from an equidistant grid $\{0.03, 0.06, \cdots, 0.30\}$ such that the loss $\widehat{Q}(\widehat{\beta})$ is minimum. The simulation results based on 500 replicates with a sample of $n = 50$ draws are shown in Table 3.1. We report the elementary descriptive statistics, mean, median, standard deviation, mean squared error and the absolute estimation error (aee) that is defined as $|\beta_{0,2} - \widehat{\beta}_2| + |\beta_{0,3} - \widehat{\beta}_3|$. The results obtained from the EFM approach proposed by Cui *et al.* (2001), adjusted by a final Fisher scoring step, as could be found in the codes kindly provided by the authors, are also reported. Moreover we report the benchmark results obtained by nonlinear least squares method (NLS) in the homoscedastic case and by weighted nonlinear least squares method (WNLS) in the heteroscedastic case. With these parametric estimation approaches the conditional mean and the conditional variance are known up to the parameter β_0. The results show that our method performs well compared to EFM. It shows slightly less performance in the homoscedastic case, but slightly outperforms EFM in the heteroscedastic case. As expected, the parametric approaches are more accurate.

Table 3.1: Single-index in mean. Simulation results for the estimators of β_0 obtained from 500 replications generated using the model (V.10).

Homoscedastic case, $n = 50$					Heteroscedastic case, $n = 50$			
		NLS	Ours	EFM		WNLS	Ours	EFM
$\beta_{0,2} = 0.8$	median	0.7992	0.8025	0.7994	median	0.7987	0.8018	0.7965
	mean	0.7991	0.8030	0.7993	mean	0.7972	0.8096	0.8070
	std	0.0163	0.0607	0.0376	std	0.0168	0.0935	0.1244
	mse	0.0002	0.0014	0.0036	mse	0.0003	0.0088	0.0155
$\beta_{0,3} = 0.5$	median	0.5001	0.4981	0.5000	median	0.4987	0.4996	0.4995
	mean	0.4996	0.5025	0.4997	mean	0.4982	0.5057	0.5072
	std	0.0164	0.0477	0.0390	std	0.0117	0.0712	0.0988
	mse	0.0002	0.0022	0.0008	mse	0.0001	0.0050	0.0098
aee		0.0252	0.0836	0.0532	aee	0.0204	0.1251	0.1664

Next we consider a third setup inspired by the empirical study presented by Ma & Zhu (2013). The law of the six covariates vector $X = (X_{(1)}, \cdots, X_{(6)})^\top$ is constructed as follows:

1 $X_{(1)}, X_{(2)}, e_1$ and e_2 are independent standard normal random variables;

2 $X_{(3)} = 0.2X_{(1)} + 0.2(X_{(2)} + 2)^2 + 0.2e_1$ and $X_{(4)} = 0.1 + 0.1(X_{(1)} + X_{(2)}) + 0.3(X_{(1)} + 1.5)^2 + 0.2e_2$;

3 given $X_{(1)}$ and $X_{(2)}$, then generate $X_{(5)}$ and $X_{(6)}$ independently as Bernoulli variables with respective success probabilities $\exp(X_1)/\{1 + \exp(X_1)\}$ and $\exp(X_2)/\{1 + \exp(X_2)\}$.

Let $\beta_0 = (1.3, -1.3, 1, -0.5, 0.5, -0.5)^\top/1.3$. The response Y is obtained as

$$Y = \sin(2X^\top\beta_0) + 2\exp(X^\top\beta_0) + \varepsilon, \tag{V.11}$$

where $\varepsilon \sim N(0, \log\{2 + (X^\top\beta_0)^2\})$. Again, we compare our method with EFM and WNLS. The results presented in Table 3.2 are obtained from 500 replications with samples of $n = 100$. Again, the bandwidth h is chosen by minimizing the loss $\widehat{Q}_n(\widehat{\beta})$ over the grid $\{0.05, 0.1, 0.15\}$. The EFM approach produces very poor results, while our method provides accurate estimates, with performance close to that of the WNLS estimates. The very good accuracy of the WNLS estimators could be explained by the construction of the setup with yields a value of the signal-to-noise ratio close to 2700.

V.2 Simulation experiments with single-index in law models

Three setups with responses having a single-index conditional law are considered. First,

$$Y = X^\top\beta_0 + \varepsilon, \tag{V.12}$$

with X a trivariate normal random vector with mean zeros, standard deviations equal to 1 and pairwise correlations equal to 0.2, and a Cauchy distribution error term. Next, following Ma & Zhu (2013), we consider

$$Y = \sin(2X^\top\beta_0) + 2\exp(X^\top\beta_0) + \varepsilon \tag{V.13}$$

Table 3.2: Single-index in mean. Simulation results for the estimators of β_0 obtained from 500 replications from the model (V.11).

		mean	std	median	mse
$\beta_2 = -1$	WNLS	-1	0.004	-1	$1.6092 * 10^{-5}$
	Ours	-1.012	0.038	-1.012	0.0016
	EFM	-3.955	3.723	-4.205	22.5622
$\beta_3 \approx 0.769$	WNLS	0.769	0.006	0.769	$3.4043 * 10^{-5}$
	Ours	0.777	0.033	0.776	0.0011
	EFM	3.181	3.003	3.196	0.148091
$\beta_4 \approx -0.385$	WNLS	-0.384	0.003	-0.385	$8.0883 * 10^{-6}$
	Ours	-0.380	0.012	-0.380	0.0001
	EFM	-1.223	2.249	-0.661	5.7518
$\beta_5 \approx 0.385$	WNLS	0.385	0.007	0.385	$5.0449 * 10^{-5}$
	Ours	0.390	0.017	0.390	0.0003
	EFM	1.193	1.295	1.373	2.3268
$\beta_6 \approx -0.385$	WNLS	-0.385	0.007	-0.384	$5.0710 * 10^{-5}$
	Ours	-0.388	0.015	-0.387	0.0002
	EFM	-1.355	1.233	-1.455	2.4593
aee	WNLS:0.0213		Ours:0.0923		EFM:10.0191

and

$$Y = \sin(2X^\top \beta_0) + 2\exp(X^\top \beta_0) + \sqrt{\log(2 + X^\top \beta_0)}\,\varepsilon, \tag{V.14}$$

where the error ε has a normal random variable and the vector of covariates $X = (X_{(1)}, X_{(2)}, X_{(3)})^\top$ is generated as follows:

1 $X_{(1)}$ and e_1 are independent standard normal random variables;

2 $X_{(2)} = 0.3 + 0.2X_{(1)} + 0.1(X_{(1)} + 1.5)^2 - 0.3e_1^2$.

3 given $X_{(1)}$ and $X_{(2)}$, $X_{(3)}$ is a Bernoulli variable with probability $\exp(X_{(1)})/\{1+\exp(X_{(1)})\}$.

In all the three examples (V.12) to (V.14), the real parameter value is $\beta_0 = (1, 0.8, -0.5)^\top$. The simulation results are based on 200 replicates with samples of $n = 50$ independent draws are reported in Table 3.3. Our method is compared with the maximum likelihood estimation (MLE), the method (PLISE) in the Chiang & Huang (2012) and the method proposed in Ma & Zhu (2013) denoted as Eff. The bandwidth h is selected as the minimum of the loss $\widehat{Q}(\widehat{\beta})$ on the grid $\{0.01, 0.02, \ldots, 0.05\}$. In the conditional Cauchy responses cases, our method performs much better than PLISE and slightly better than Eff. The bad behavior of

the MLE is likely connected to the multiple local maxima of a Cauchy likelihood, a well known problem in the classical statistics. See, for instance, Reeds (1985). In the conditional gaussian examples, our methods seems to outperform the semiparametric competitors with respect to almost all the indicators we provide (mean, median, standard deviation and absolute estimation error).

V.3 Bootstrap confidence interval

Next, we use the idea described in section IV to build confidence intervals for the components of β in the models (V.10) and (V.12). We consider 200 samples of $n = 50$ independent draws and for each sample we generated 199 independent random samples $\xi_1, \ldots, \xi_n$ for an exponential law with parameter equal to 1, and computed the criteria $\widehat{Q}^*(\beta)$. The 90% and 95% confidence intervals obtained with the optimal values $\widehat{\beta}^*$ are presented in Table 3.4. The level is quite accurate and the intervals have reasonable length, indicating that our simulation based procedure for building confidence intervals is quite effective.

V.4 Real data applications

The investigation of the finite sample performances of our semiparametric approach is completed by two applications using real data.

The first example is the New York air quality data set. See Chambers *et al.* (1983). It contains the measurements of daily ozone concentration (*ozone*), wind speed (*wind*), daily maximum temperature (*temp*), and solar radiation level (*solar*) on 111 successive days from May to September 1973 in New York metropolitan area. The response variable is *ozone*, with empirical mean 42.0991 and empirical variance 1107.29. Yu & Ruppert (2002), Anestis *et al.* (2004), Kong & Xia (2007) considered a single-index mean regression model for this data set, while and Chiang & Huang(2012) fitted a single-index in law model. Here we consider the covariate vector X with components the variables $wind, temp, wind^2, solar^2, wind * temp,$ and $temp * solar$ and we consider the single-index in mean assumption. The coefficient of *wind* is set equal to 1. The single-index assumption was checked using

Table 3.3: Single-index in conditional law. Simulation results from 200 replications with the models (V.12) to (V.14).

Model (V.12), conditional Cauchy response, $n = 50$					
		MLE	PLISE	Eff	Ours
$\beta_2 = 0.8$	median	0.7924	0.8245	0.7956	0.8020
	mean	0.9063	0.9647	0.7971	0.8082
	std	0.5872	0.7635	0.1268	0.1113
	mse	0.3544	0.6072	0.0160	0.0123
$\beta_3 = -0.5$	median	-0.5169	-0.5351	-0.5426	-0.5405
	mean	-0.5621	-0.5801	-0.5414	-0.5452
	std	0.3459	0.3883	0.0807	0.0879
	mse	0.0.1229	0.1564	0.0082	0.0097
aee		0.6119	0.7295	0.1700	0.1585
Model (V.13), conditional gaussian, homoscedastic response, $n = 50$					
		MLE	PLISE	Eff	Ours
$\beta_2 = 0.8$	median	0.8020	0.8156	0.8116	0.7977
	mean	0.8002	0.8193	0.8161	0.7989
	std	0.0231	0.1634	0.1598	0.1158
	mse	0.0005	0.0269	0.0256	0.0133
$\beta_3 = -0.5$	median	-0.5003	-0.5015	-0.5116	-0.5105
	mean	-0.4998	-0.4998	-0.4998	-0.5166
	std	0.0094	0.0713	0.0982	0.0720
	mse	0.00008	0.0050	0.0096	0.0054
aee		0.0243	0.1797	0.1944	0.1411
Model (V.14), conditional gaussian, heteroscedastic response, $n = 50$					
		MLE	PLISE	Eff	Ours
$\beta_2 = 0.8$	median	0.8010	0.8131	0.8197	0.7922
	mean	0.8007	0.8203	0.8278	0.7887
	std	0.0250	0.1752	0.2171	0.1072
	mse	0.0006	0.0309	0.0477	0.0115
$\beta_3 = -0.5$	median	-0.5102	-0.5023	-0.4974	-0.5004
	mean	-0.5000	-0.5022	-0.4952	-0.5114
	std	0.0109	0.0734	0.1022	0.0676
	mse	0.0001	0.0053	0.0104	0.0046
aee		0.0272	0.1894	0.2229	0.1373

Table 3.4: Empirical level and empirical length for the componentwise bootstrap confidence intervals in the models (V.10) and (V.12): sample size $n = 50$ and 199 bootstrap samples

Model (V.10) with $N(0, 0.25)$ errors				
	90% bootstrap CI		95% bootstrap CI	
	length	level	length	level
$\beta_2 = 0.8$	0.1592	92	0.2054	96.5
$\beta_3 = 0.5$	0.1785	89	0.2301	96
Model (V.12) with Cauchy errors				
	90% bootstrap CI		95% bootstrap CI	
	length	level	length	level
$\beta_2 = 0.8$	0.2822	89.5	0.3713	96
$\beta_3 = -0.5$	0.1923	90	0.2538	96

the test proposed by Maistre & Patilea (2014) with bootstrap critical values and the $p-$value was 0.403. The estimate of the direction β and the componentwise confidence intervals are given in Table 3.5. The plot of the estimated link function is provided in Figure 3.1. The mean absolute deviation is

$$\frac{1}{111}\sum_{i=1}^{111}\left|ozone_i - \widehat{\mathbb{E}}[ozone_i|X_i^\top\widehat{\beta}]\right| = 17.9248.$$

To estimate the parameter β we select the bandwidth by minimization of the loss $Q(\widehat{\beta})$ on a grid $\{0.01, 0.02, \cdots, 0.09\}$ Given the estimate $\widehat{\beta}$, we build the adjusted values by univariate smoothing of the response given $X_i\widehat{\beta}$ with a bandwidth selected by least-squares cross-validation.

Table 3.5: The estimator $\widehat{\beta}$ and the componentwise bootstrap confidence intervals (BCI) (levels 0.9 and 0.95) for New York air quality data: single-index mean regression model.

Variable	Coefficient estimate	0.9 BCI	0.95 BCI
$temp$	-6.0144	(-6.3278, -5.7520)	(-6.4813, -5.6591)
$wind^2$	-3.1942	(-3.2430, -2.9008)	(-3.2914, -2.8394)
$solar^2$	-1.3832	(-1.5161, -1.1607)	(-1.6557, -1.0798)
$wind * temp$	-0.5791	(-0.7472, -0.1098)	(-0.8292, -0.0554)
$temp * solar$	1.5339	(1.3256, 1.7995)	(1.2714, 1.8581)

The second real data example illustrates the single-index in law model. We consider the data on the employees' salaries in the Fifth National Bank of Springfield,

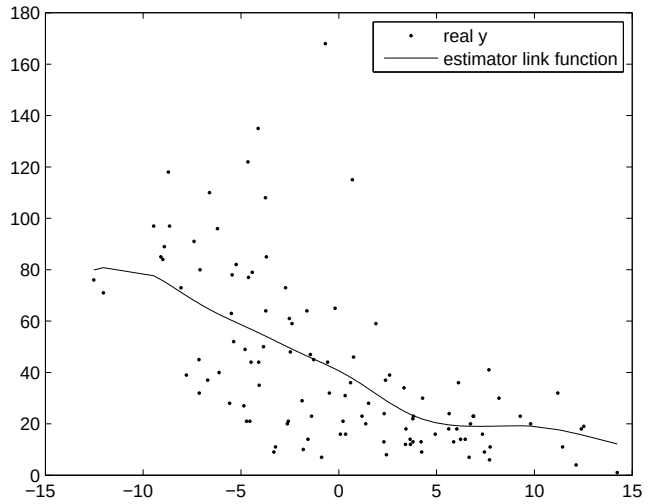


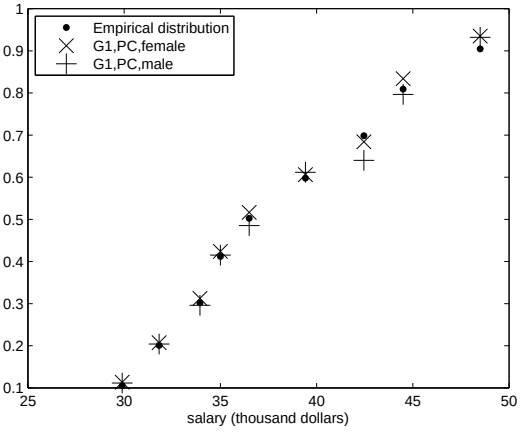
Figure 3.1: The estimated link function for New York air quality data set.

see Albright *et al.* (1999). There are 208 observations in the data set and every observation contains 8 variables: *Education* (a categorical variable with 5 education levels), *Grade* (a categorical variable with 6 job level), *Year1* (years of work experience at Fifth National), *Age* (employee's current age), *Year2* (years of work experience at another bank prior to working at Fifth National), *Gender* ('female'=1, 'male'=0), *PC Job* (a categorical variable depending on whether the job is computer related, 'yes'=1, 'no'=0), *Salary* (annual salary, the response variable). Like in Fan & Peng (2004), we delete the observations with *Age* over 60 or working experience $Year1 + Year2$ over 30 and this results in a subsample of 199 observations. Next, following Ma & Zhu (2013), we drop the variable *Education*, set the coefficient of *Grade* equal to 1 and let *Grade* take values from 1 to 6. The single-index assumption for the conditional law was checked using the test proposed by Maistre & Patilea (2014) with asymptotic critical values and the $p-$value was 0.166. The estimator $\widehat{\beta}$ obtained by our approach is reported in Table 3.6, together with the bootstrap confidence intervals. On contrary to the results reported by Ma & Zhu (2013), we found a significant negative coefficient for *Gender*. This could be explained by the negative correlation between *Gender* and *Grade*. For instance, there is no female with *Grade*=6 in our working sample. In Figures 3.2, 3.3 and 3.4, we show the estimates of the values of the conditional distribution functions and of the empirical

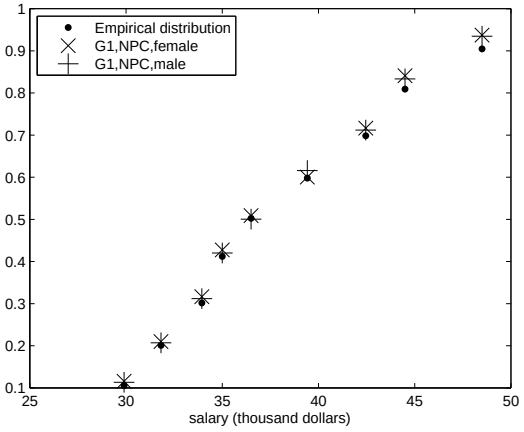
distribution function for ten values of the response. The values of response were determined as the empirical deciles of the observed responses. We plot the kernel estimates of the conditional distribution functions for three different job levels (*Grade*=1,3 and 5, respectively). For each of the three job levels, we compute the estimates of the conditional distribution given $X = x$ for four different values of x. These values x correspond to all the possible outcomes for the variables *Gender* and *PC Job*. The components corresponding to the covariates *Year1*, *Age*, *Year2*, are set equal to the average values of the subsamples obtained with the given job level, and with *PCJob=1* or 0, for each gender. For each value of the conditional distribution function estimated by kernel smoothing, we selected the bandwidth by least-squares cross-validation. In most cases, the figures reveal little difference between the distribution functions for female and male, which confirms the usual conclusion that could be found in the literature, *i.e.,* there is not evidence in the Fifth National data set that the female employees are discriminated. See, for instance, Fan & Peng (2004).

Table 3.6: The estimator $\widehat{\beta}$ and the componentwise bootstrap confidence intervals (BCI) (levels 0.9 and 0.95) for Fifth National Bank of Springfield salary data: single-index in law model.

Variable		Estimation	0.9 BCI	0.95 BCI
$Year1$	Ours	0.513	(0.485,0.562)	(0.478 ,0.591)
	Eff	0.477	(0.442, 0.511)	(0.435, 0.518)
Age	Ours	0.727	(0.689, 0.753)	(0.675, 0.775)
	Eff	0.265	(0.214, 0.315)	(0.204, 0.325)
$Year2$	Ours	0.086	(0.0487, 0.103)	(0.036 , 0.114)
	Eff	0.024	(-0.025, 0.073)	(-0.034, 0.082)
$Gender$	Ours	-0.783	(-0.839 ,-0.657)	(-0.867 ,-0.633)
	Eff	0.050	(-0.010, 0.110)	(-0.022, 0.122)
$PCJob$	Ours	0.689	(0.675,0.934)	(0.652, 0.998)
	Eff	0.146	(0.095, 0.1968)	(0.085, 0.206)

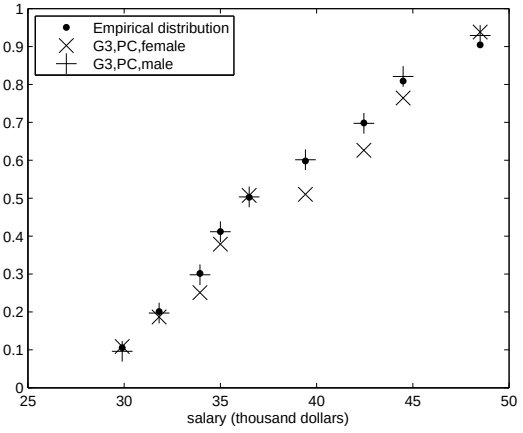


(a) the job is computer related

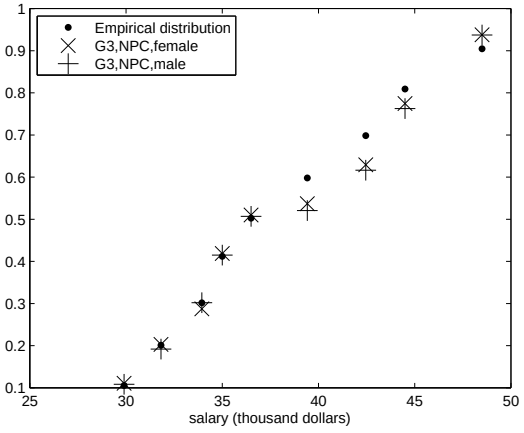


(b) the job is not computer related

Figure 3.2: The conditional distribution function of the Fifth National Bank salary data set for $Grade = 1$: $Year1$, Age, $Year2$ take the sample mean value given $Grade = 1$ and the values of *Gender* and *PC Job*

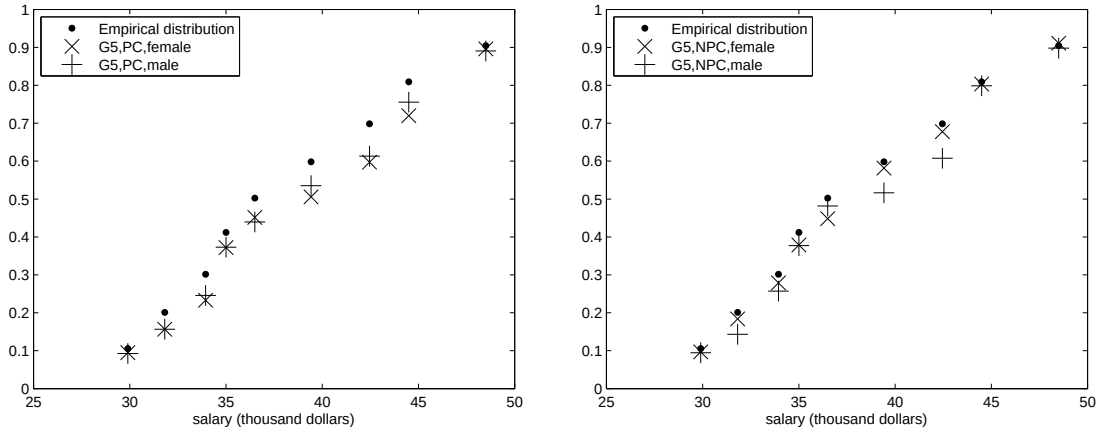


(a) the job is computer related



(b) the job is not computer related

Figure 3.3: The same plots as in Figure 3.2 in the case *Grade=3*.



(a) the job is computer related

(b) the job is not computer related

Figure 3.4: The same plots as in Figure 3.2 in the case *Grade=5*.

VI References

1. Albright S.C., Winston W.L. & Zappe, C.J. (1999). *Data Analysis and Decision Making with Microsoft Excel*. Duxbury Press, Pacific Grove, California.

2. Anestis A., Gérard G. & Ian W.M. (2004). Bayesian estimation in single-index models. *Statistica Sinica* **14**, 1147-1164.

3. Carroll R.J., Fan J.Q. & Wand M.P. (1997). Generalized Partially Linear Single-Index Models. *J. Amer. Statist. Assoc.* **92**, 477–489.

4. Chiang C.-T. & Huang M.-Y. (2012). New estimation and inference procedures for a single-index conditional distribution model. *J. Multivariate Anal.* **111**, 271–285.

5. Chambers J.M., Cleveland W.S., Kleiner, B. & Tukey P.A. (1983). *Graphical Methods for Data Analysis.* Belmont, CA: Wadsworth.

6. Cui X., Härdle W., & Zhu L. (2011). The EFM approach for single-index models. *Ann. Statist.* **39**, 1658-1688.

7. Delecroix M., Härdle W. & Hristache M. (2003). Efficient estimation in conditional single-index regression. *J. Multivariate Anal.* **86**, 213–226.

8. Delecroix, M., Hristache, M. & Patilea, V. (2006). On semiparamet-

ric $M-$estimation in single-index regression. *J. Statist. Plan. Inference* **136**, 730–769.

9. Fan J. & Peng H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32**, 928–961.
10. Hall P. & Yao Q. (2005). Approximating conditional distribution functions using dimension reduction. *Ann. Statist.* **33**, 1404–1421.
11. Härdle W., Hall P. & Ichimura H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21**, 157–178.
12. Hristache M., Juditsky A. & Spokoiny V. (2001). Direct estimation of the index coefficient in a single-index model. *Ann. Statist.* **29**, 595–917.
13. Horowitz J.L. (2009). *Semiparametric and nonparametric methods in econometrics.* Springer-Verlag. New York
14. Ichimura H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single index models. *Journal of Econometrics* **58**, 71–120.
15. Jin Z., Ying Z. & Wei L.J. (2001). A simple resampling method by perturing the minimand. *Biometrika* **88**, 381–390.
16. Klein R.W. & Spady R.H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica* **61**, 387–421.
17. Kong E. & Xia Y. (2007). Variable selection for the single-index model. *Biometrika* **94**, 217-229.
18. Kong E. & Xia Y. (2012). A single-index quantile regression model and its estimation. *Econometric Theory* **28**, 730–768.
19. Lavergne P. & Patilea V. (2013). Smooth minimum distance estimation and testing with conditional estimating equations: Uniform in bandwidth theory. *J. Econometrics* **177**, 47–59.
20. Li W. & Patilea V. (2015). A dimension reduction approach for conditional Kaplan-Meier estimators. CREST-Ensai working paper.
21. Liang H., Liu X., Li R. & Tsai C.L. (2010). Estimation and testing for partially linear single-index models. *Ann Statist.* **38**, 3811–3836.

22. Ma Y. & Zhu L. (2013). Efficient Estimation in Sufficient Dimension Reduction. *Ann. Statist.* **41**, 250–268.

23. Maistre S. & Patilea V. (2014). Nonparametric model checks of single-index assumptions. arXiv:1410.4934[math.ST]

24. Powell J.L., Stock J.H. & Stoker T.M. (1989). Semiparametric estimation of index coefficients. *Econometrica* **51**,1403–1430.

25. Reeds J.A. (1985). Asymptotic Number of Roots of Cauchy Location Likelihood Equations. *Ann. Statist.* **13**, 775–784.

26. van der Vaart A.D. (1998). *Asymptotic Statistics.* Cambridge University Press.

27. van der Vaart A.D. & Wellner, J.A. (2011). A local maximal inequality under uniform entropy. *Electron. J. Statist.* **5**, 192–203.

28. Yin X. & Cook R.D. (2002). Dimension reduction for the conditional $k-$th moment in regression. *J. R. Statist. Soc. B* **64**, 159–175.

29. Yu Y. & Ruppert D. (2002). Penalized spline estimation for partially linear single index models. *J. Amer. Statist. Assoc.* **97**, 1042–1054.

30. Zhang J., Feng Z. & Xu P. (2015). Estimating the conditional single-index error distribution with a partial linear mean regression. *Test* **24**, 61–83.

VII Appendix

Assumption VII.1. *1. The observations (Y_i, X_i), $1 \leq i \leq n$, are independent copies of $(Y, X) \in \mathbb{R}^d \times \mathbb{R}^p$.*

2. The parameter set is $\mathcal{B} = \{1\} \times \mathcal{B}'$ and $\mathcal{B}' \subset \mathbb{R}^{p-1}$ is a compact set. The vector $\beta_0 \in \mathcal{B}$ satisfying the condition (II.6) is the unique element $\mathcal{B}$. For any $\beta \in \mathcal{B}$ the random variable $X^\top \beta$ has a density f_β such that $\sup_{\beta \in \mathcal{B}} \sup_{z \in \mathbb{R}} f_\beta(z) < \infty$.

3. We have $\sup_{u \in \mathcal{U}} \sup_{\beta \in \mathcal{B}} \sup_{z \in \mathbb{R}} |\mathbb{E}(T_u(Y) \mid X^\top \beta = z)| f_\beta(z) < \infty$ and

$$\lim_{\delta \to 0} \sup_{\beta \in \mathcal{B}} \sup_{z \in \mathbb{R}} |f_\beta(z + \delta) - f_\beta(z)| = 0,$$

$$\lim_{\delta \to 0} \sup_{u \in \mathcal{U}} \sup_{\beta \in \mathcal{B}} \sup_{z \in \mathbb{R}} \Big| \Big(\mathbb{E}[T_u(Y) \,|\, X^\top \beta = \cdot\,] f_\beta \Big)(z+\delta) - \Big(\mathbb{E}[T_u(Y) \,|\, X^\top \beta = \cdot\,] f_\beta \Big)(z) \Big| = 0.$$

4. The family of transformations $\{T_u(\cdot) : u \in \mathcal{U}\}$ is a VC-class (or Euclidian) for an envelope with finite moment of order $4 + \rho$ for some $\rho > 0$.

5. The value $\boldsymbol{\beta}_0$ is a well-separated point of minimum for $Q(\beta)$ defined in equation (II.7) with $\omega(x) = \exp(-\|x\|^2/2)$ and μ equal to the distribution F_Y of the observations Y that means, for any $\varepsilon > 0$, $\inf_{\beta \in \mathcal{B}, \|\beta - \beta_0\| \geq \varepsilon} Q(\beta) > Q(\beta_0)$.

6. The kernel $K(\cdot)$ is a univariate integrable function with bounded variation. The bandwidth h satisfies the condition $h + n^{-1}h^{-2} \to 0$.

Let us introduce some notation. Let $\widetilde{X} \in \mathbb{R}^{p-1}$ be the $(p-1)-$dimension vector of the last components of X. Below, $(\widetilde{X})_r$ (resp. $(\widetilde{X}\widetilde{X}^\top)_{rq}$) denotes the rth components (resp. the $rq-$entry) of the vector $\widetilde{X}$ (resp. matrix $\widetilde{X}\widetilde{X}^\top$). If A is a matrix with real entries, $\|A\| = \sqrt{trace(A^\top A)}$. In the following, where ∂_z (resp. ∂^2_{zz}) denotes the first (resp. second) order derivative with respect to z.

Assumption VII.2. *1. There exists a positive number a such that*

$$\mathbb{E}[\exp(a\|X\|)] < \infty.$$

2. The subvector β_0 built with the last $(p-1)$ components belong to the interior of $\mathcal{B}'$, where $\mathcal{B} = \{1\} \times \mathcal{B}'$.

3.

$$\sup_{z\in\mathbb{R}} \mathbb{E}\left[\|\widetilde{X}\|^4 \mid X^\top\beta_0 = z\right] f_{\beta_0}(z) < \infty \tag{VII.1}$$

4. The functions $z \mapsto \mathbb{E}\left[T_u(Y) \mid X^\top\beta_0 = z\right]$, $u \in \mathcal{U}$, *and*

$$z \mapsto f_{\beta_0}(z), \quad z \mapsto \mathbb{E}[(\widetilde{X})_r \mid X^\top\beta_0 = z] \quad \textit{and} \quad z \mapsto \mathbb{E}[(\widetilde{X}\widetilde{X}^\top)_{rq} \mid X^\top\beta_0 = z]$$

are four times continuously differentiable and the derivatives up to order four are bounded. The fourth order derivative are Lipschitz functions. The Lipschitz constant is independent of u *in the case of the four order derivative of* $\mathbb{E}\left[T_u(Y) \mid X^\top\beta_0 = z\right]$.

5. Let A *be the set of values* $u \in \mathcal{U}$ *such that*

$$Var\left[\left(\widetilde{X} - \mathbb{E}\left[\widetilde{X} \mid X^\top\beta_0\right]\right)\partial_z\left\{\mathbb{E}[T_u(Y) \mid \cdot]\right\}(X^\top\beta_0)\right] \tag{VII.2}$$

is positive definite. Then $F_Y(A) > 0$.

6. Let $z \mapsto \lambda_\beta(z;u)$ *denote any of the four functions at point (4) above, considered for each* $\beta \in \mathcal{B}$, *and their derivatives up to the second order. Then, the family of functions* $\{\lambda_\beta(\cdot) : \beta \in \mathcal{B}, u \in \mathcal{U}\}$ *is a VC-class (or Euclidian) for an envelope having a finite moment of order 8. Moreover, for any sequence* $b_n \to 0$,

$$\sup_{\|\beta-\beta_0\|\le b_n} \sup_{z\in\mathbb{R}} \sup_{z\in\mathbb{R}} |\lambda_\beta(z;u) - \lambda_{\beta_0}(z;u)| \to 0.$$

7. The kernel $K(\cdot)$ *is a symmetric and twice continuously differentiable univariate density with the second order derivative with bounded variation. Moreover, for* $\kappa = 1, 2$, $\int_{\mathbb{R}} |K^{(\kappa)}(u)|du < \infty$, *where* $K^{(\kappa)}(\cdot)$ *denotes the* κ*th derivative of* $K(\cdot)$.

8. $nh^4 \to 0$ *and* $nh^{3+a} \to \infty$ *for some* $a \in (0,1)$.

VII.1 Proofs

Proof of Lemma II.1. Let $\mathcal{F}[\omega]v) = \int_{\mathbb{R}^p} e^{-2\pi i x^\top v}\omega(x)dx$, $u \in \mathbb{R}^p$, denote the Fourier Transform of $\omega(\cdot)$. If $\mathcal{F}[\omega]$ is integrable, by the Inverse Fourier Transform formula and Fubini Theorem, we can write

$$\begin{aligned} Q(\beta) &= \int_{\mathcal{U}} \mathbb{E}\left[\omega(X_1 - X_2) g_u(Y_1, X_1^\top\beta;\beta)^\top g_u(Y_2, X_2^\top\beta;\beta)\right] d\mu(u) \\ &= \int_{\mathcal{U}} \mathbb{E}\left[g_u(Y_1, X_1^\top\beta;\beta)^\top g_u(Y_2, X_2^\top\beta;\beta) \int_{\mathbb{R}^p} e^{2\pi i (X_1 - X_2)^\top v}\mathcal{F}[\omega](v)dv\right] d\mu(u) \\ &= \int_{\mathcal{U}}\int_{\mathbb{R}^p} \left\|\mathbb{E}\left[\mathbb{E}\left[g_u(Y, X^\top\beta;\beta) \mid X\right] e^{2\pi i X^\top v}\right]\right\|^2 \mathcal{F}[\omega](v)dv d\mu(u). \end{aligned}$$

By the fact that $\mathcal{F}[\omega]$ is positive, $Q(\beta) \geq 0$, $\forall \beta \in \mathcal{B}$. Using also the uniqueness of the Fourier Transform, one can deduce that

$$Q(\beta) = 0 \quad \Leftrightarrow \quad \mathbb{E}\left[g_u(Y, X^\top\beta;\beta) \mid X\right] = 0 \quad \text{almost surely, for } \mu-\text{almost all } u \in \mathcal{U}.$$

The conclusion of the lemma follows from definition of the functions g_u and the transformations T_u (that is, $T_u(y) = y$, $\forall u$, or $T_u(y) = \mathbf{1}\{y \leq u\}$, $u \in \mathbb{R}^p$). □

Proof of Proposition 3.1. The proof of the asymptotic normality is a particular case of the asymptotic normality result of Li & Patilea (2015), and hence will be omitted. Concerning the consistency, by the Assumption VII.1-5, β_0 is a well-separated point of minimum for $Q(\beta)$. Thus, it suffices to prove that

$$\sup_{\beta\in\mathcal{B}} \left|\widehat{Q}_n(\beta) - Q(\beta)\right| = o_{\mathbb{P}}(1). \tag{VII.3}$$

See, for instance, Theorem 5.7 of van der Vaart (1998). For this purpose, let us simplify notation and write $\widehat{g}_{u,i}(\beta)$ instead of $\widehat{g}_u(Y_i, X_i^\top\beta;\beta)$. By Lemma VII.1,

$$\sup_{\beta\in\mathcal{B}}\sup_{u\in\mathcal{U}} \left| \frac{1}{n^2}\sum_{i,j=1}^n \widehat{g}_u(Y_i, X_i^\top\beta;\beta)^\top \widehat{g}_u(Y_j, X_j^\top\beta;\beta)\omega_{ij} \right.$$
$$\left. - \frac{1}{n^2}\sum_{i,j=1}^n g_u(Y_i, X_i^\top\beta;\beta)^\top g_u(Y_j, X_j^\top\beta;\beta)\omega_{ij} \right| = o_{\mathbb{P}}(1).$$

Next, by the uniform law of large numbers for Glivenko-Cantelli classes of functions (see, for instance, van der Vaart (1998), Theorem 19.4), we deduce

$$\sup_{\beta\in\mathcal{B}}\sup_{u\in\mathcal{U}}\left|\frac{1}{n^2}\sum_{i,j=1}^{n}\widehat{g}_u(Y_i,X_i^\top\beta;\beta)^\top\widehat{g}_u(Y_j,X_j^\top\beta;\beta)\omega_{ij} - \mathbb{E}\left[g_u(Y_1,X_1^\top\beta;\beta)^\top g_u(Y_2,X_2^\top\beta;\beta)\omega_{12}\right]\right| = o_{\mathbb{P}}(1).$$

From this, it follows

$$\sup_{\beta\in\mathcal{B}}\left|\widehat{Q}(\beta) - \int_{\mathcal{U}}\mathbb{E}\left[g_u(Y_1,X_1^\top\beta;\beta)^\top g_u(Y_2,X_2^\top\beta;\beta)\omega_{12}\right]d\mu_n(u)\right| = o_{\mathbb{P}}(1).$$

Next, by the uniform law of large numbers for Glivenko-Cantelli classes of functions,

$$\sup_{\beta\in\mathcal{B}}\left|\int_{\mathcal{U}}\mathbb{E}\left[g_u(Y_1,X_1^\top\beta;\beta)^\top g_u(Y_2,X_2^\top\beta;\beta)\omega_{12}\right]d\mu_n(u) - Q(\beta)\right| = o_{\mathbb{P}}(1).$$

Gathering facts, deduce that the uniform convergence in equation (VII.3) holds true, and thus $\widehat{\beta}$ is consistent in probability. □

Lemma VII.1. *Under the Assumption VII.1*

$$\sup_{1\leq i\leq n}\sup_{u\in\mathcal{U}}\sup_{\beta\in\mathcal{B}}\left\|\widehat{g}_u(Y_i,X_i^\top\beta;\beta) - g_u(Y_i,X_i^\top\beta;\beta)\right\| = o_{\mathbb{P}}(1).$$

Proof of Lemma VII.1. The result follows from the following two properties:

$$\sup_{1\leq i\leq n}\sup_{\beta\in\mathcal{B}}\left\|\widehat{f_\beta(X_i^\top\beta)} - f_\beta(X_i^\top\beta)\right\| = o_{\mathbb{P}}(1)$$

and

$$\sup_{1\leq i\leq n}\sup_{u\in\mathcal{U}}\sup_{\beta\in\mathcal{B}}\left\|\mathbb{E}[T_u(Y_i)\mid\widehat{X_i^\top\beta}]f_\beta(X_i^\top\beta) - \mathbb{E}[T_u(Y_i)\mid X_i^\top\beta]f_\beta(X_i^\top\beta)\right\| = o_{\mathbb{P}}(1). \tag{VII.4}$$

Since the first property is a particular case of the second one, we only provide the justification for the equation (VII.4). The latter property is a direct consequence of

the following statements:

$$\sup_{z\in\mathbb{R}}\sup_{u\in\mathcal{U}}\sup_{\beta\in\mathcal{B}}\left|\mathbb{E}\left[T_u(Y)K((X^\top\beta-z)/h)\right]-\mathbb{E}\left[T_u(Y)\mid X^\top\beta=z\right]f_\beta(z)\right|=o(1) \tag{VII.5}$$

and

$$\sup_{z\in\mathbb{R}}\sup_{u\in\mathcal{U}}\left|\frac{1}{nh}\sum_{k=1}^{n}T_u(Y_k)K((X_k^\top\beta-z)/h)-\mathbb{E}\left[T_u(Y)K((X^\top\beta-z)/h)\right]\right|=o_{\mathbb{P}}(1). \tag{VII.6}$$

The statement (VII.5) follows by a standard change of variables and the Assumption VII.1-3. For the uniform convergence in equation (VII.6), it suffices, for instance, to use the Maximal Inequality of van de Vaart & Wellner (2011), Theorem 3.1. In that result it suffices to take p sufficiently large such that $(4p-2)/(p-1)\leq 4+\rho$, with ρ from Assumption VII.1-4, and apply the maximal inequality with $\delta=h^{1/2}$. Deduce that

$$\begin{aligned}\sup_{z\in\mathbb{R}}\sup_{u\in\mathcal{U}}&\left|\frac{1}{nh}\sum_{k=1}^{n}T_u(Y_k)K((X_k^\top\beta-z)/h)-\mathbb{E}\left[T_u(Y)K((X^\top\beta-z)/h)\right]\right|\\&=O_{\mathbb{P}}(n^{-1/2}h^{-1/2}\log^{1/2}n)=o_{\mathbb{P}}(1).\end{aligned}$$

Now the proof is complete. □

4

A dimension reduction approach for conditional Kaplan-Meier estimators

I Introduction

Survival data, also called duration or time to event data, are often incomplete due to the presence of censoring. In such cases, the model has to take into account the censoring mechanism in order to avoid substantial biases. Several types of models and inference approaches have been developed, among which the proportional hazards model and the partial likelihood estimation method of Cox (1972) with right-censored data are perhaps the most famous. In a more flexible, nonparametric regression perspective, the conditional distribution of the response given the covariates in the presence of right censoring is usually estimated by the Beran (1981)'s estimator, also called the conditional Kaplan-Meier estimator. See also Dabrowska (1989, 1992) who pointed out that Beran's estimator is a smooth functional of the kernel regression estimator of the conditional law of the observations. An obvious drawback of the nonparametric approach is the curse of dimensionality that one faces with multi-dimensional covariates.

A common remedy to the curse of dimensionality consists in considering a dimension reduction device before applying the nonparametric smoother. Semiparametric index regression is a common example of dimension reduction approach that has been already used to estimate conditional laws with complete data, see for instance Hall & Yao (2005), Chiang & Huang (2012), Ma & Zhu (2013) and Lee *et al.* (2013). Index regression is also the idea we follow in this paper where the responses could

be censored. Semiparametric index estimation of a conditional law in the presence of right censoring was considered recently by Xia *et al.* (2010), Bouziz & Lopez (2010) and Strzalkowska-Kominiak & Cao (2013). However, the existing approaches impose the dimension reduction directly on the conditional law of the lifetime of interest. In general, this results in rather complicated procedures.

In this chapter we show that there is an alternative, easier way to introduce the model assumptions. For this we recall Dabrowska's remark that in many situations, the quantities of interest in survival analysis are smooth, closed-form expression functionals of the law of the observations. This is, for instance, the case for the conditional law of the lifetime of interest under random right censoring, but also for other quantites, as for instance the conditional probability of being cured in cure survival models. Keeping this in mind, we propose to impose the dimension reduction index hypothesis directly on the conditional law of the observed variables, there are typically the possibly censored lifetime and the indicator for the presence of censoring. Next, we apply the smooth functionals to the estimator of the conditional law of the observations. This results in semiparametric estimators of the quantites of interest that avoid the curse of dimensionality. The new methodology allows to test the dimension reduction assumption and extends to other dimension reduction methods. Moreover, it can be applied to more general censoring mechanisms and closed-form expression functionals.

The paper is organized as follows. In section II we recall the general framework of right-censored data. In section III we introduce the new semiparametric index-regression approach and we focus on the single-index case. We also reconsider the general construction of Kaplan-Meier functionals in the context of the new dimension reduction idea. We end section III with a discussion on the existing single-index approaches. In particular, we shed new light on these approaches and show that our framework is not more restrictive. In section IV we propose a general semiparametric estimation method for the index and we prove the $\sqrt{n}-$normality of the index estimator. The asymptotic result is derived under mild conditions, in particular the covariates need not to be bounded or absolutely continuous, and no trimming is required. Since the estimation procedure is designed in the space of the obser-

vations, the $\sqrt{n}-$normality result remains valid even when the usual identification assumption used in the survival analysis with covariates, that is the censoring is noninformative given the covariates, fails. We end section section IV with some proposals for building confidence intervals for the index vector and for testing the dimension reduction assumption against general nonparametric alternatives. In section V we reconsider the estimation of the conditional law of the lifetime of interest using the Beran estimator with the estimated index. We derive an i.i.d. representation of the new single-index estimator. As a consequence of this representation, we introduce and prove the asymptotic normality of a new single-index estimator of the cure fraction. Our new methodology is illustrated by empirical experiments using simulated and real data. In particular, we provide a new point of view for modeling Stanford heart transplantation data. We end the paper with discussions on possible extensions of the new approach and an example of a more complicated censoring mechanism that generates a closed form expression map between the observations and the conditional law of the lifetime of interest. The assumptions and the main proofs are postponed to the Appendix. A Supplementary Material completes our work with some additional technical results.

II The framework

Let T denote the lifetime of interest that takes values in $(-\infty, \infty]$. Consider the situation where one observes independent copies of Y, δ and X, where Y is a real-valued random variable, δ is an indicator variable and X is a covariate taking values in some space $\mathcal{X}$. For the moment, we do not need to make any specific assumption about $\mathcal{X}$. The indicator variable reveals whether Y is precisely the lifetime of interest, or Y is only a random quantity smaller than T. In other words,

$$\delta = 1 \quad \text{if} \quad Y = T \qquad \text{and} \qquad \delta = 0 \quad \text{if} \quad Y < T.$$

The purpose is to estimate the law of T given X. The conditional probability of the event $\{T = \infty\}$ will be allowed to be positive.

The observations are characterized by the conditional sub-probabilities

$$\begin{aligned} H_1((-\infty, t] \mid x) &= \mathbb{P}(Y \leq t, \delta = 1 \mid X = x) \\ H_0((-\infty, t] \mid x) &= \mathbb{P}(Y \leq t, \delta = 0 \mid X = x), \qquad t \in \mathbb{R},\ x \in \mathcal{X}. \end{aligned}$$

Then the law of Y is characterized by

$$H((-\infty, t] \mid x) = \mathbb{P}(Y \leq t \mid X = x) = H_0((-\infty, t] \mid x) + H_1((-\infty, t] \mid x).$$

We suppose Y is finite, that means

$$H((-\infty, \infty) \mid x) = 1, \qquad \forall x \in \mathcal{X}.$$

The usual way to model this situation in order to estimate the conditional law T is to consider that there exists a random variable C, the right-censoring time, and

$$Y = T \wedge C, \qquad \delta = \mathbf{1}\{T \leq C\}.$$

Using suitable identification conditions, the conditional law of T given X could be expressed as a closed-form expression functional of $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$ and thus could be easily estimated by plugging in nonparametric estimates of $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$. Such an estimator of the conditional law of T given X is usually called a conditional Kaplan-Meier estimator. See Beran (1981), Dabrowska (1989), van Keilegom & Veraverbeke (1996). All these approaches suffer from the curse of dimensionality when $\mathcal{X}$ is of higher dimension than the real line. Here, we propose a dimension reduction approach for estimating $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$ and we will focus on the single-index case. This will result in a natural single-index estimator of the conditional law of T given X.

III Single-index modeling under random censoring

In this section, we introduce our general dimension reduction approach using the law of the observations and we show how it induces a dimension reduction for the quantities of interest. For the sake of simplicity, we focus on the single-index case, though the idea is clearly more general. In particular, we shed new light on the existing conditional distribution single-index models.

III.1 Dimension reduction for modeling the conditional law of the observations

Let us consider that the conditional law of (Y, δ) satisfies a dimension reduction condition. More precisely, let $\mathcal{B}$ be a parameter space, $q \geq 1$ and let

$$\lambda : \mathcal{X} \times \mathcal{B} \to \mathbb{R}^q$$

be a given map such that

$$(Y, \delta) \perp X \mid \lambda(X, \beta_0).$$

for some $\beta_0 \in \mathcal{B}$. Typically, the value of q is much smaller than the dimension of the covariates space $\mathcal{X}$. Although the dimension reduction approach is potentially more general, we will focus on the case

$$\mathcal{X} = \mathbb{R}^p, \quad \mathcal{B} \subset \mathbb{R}^p \quad \text{and} \quad \lambda(X, \beta) = X^\top \beta \in \mathbb{R}.$$

(Here and in the following, a vector is a column matrix and for any matrix A, $A^\top$ denotes its transpose.) In other words, we focus on the single-index modeling of the conditional law of the observations (Y, δ) given finite dimension covariates. Hence we suppose

$$(Y, \delta) \perp X \mid X^\top \beta_0 \tag{III.1}$$

for some $\beta_0 \in \mathcal{B}$. Similar modeling was considered, for instance, by Delecroix *et al.* (2003), Hall & Yao (2005), Chiang & Huang (2012), Ma & Zhu (2013) and Lee *et al.* (2013). However, none of the previous approaches seems appropriate in the present framework where (Y, δ) does not have a conditional density and X is not necessarily required to be absolutely continuous.

We can rewrite the condition (III.1) using the subdistributions H_0 and H_1. For any $d \in \{0, 1\}$ and $t \in \mathbb{R}$, let

$$H_{d,\beta}((-\infty, t] \mid z) = \mathbb{P}(Y \leq t, \delta = d \mid X^\top \beta = z), \qquad z \in \mathbb{R},$$

and $H_{d,\beta}(dt \mid z)$ be the associated measures. Then, condition (III.1) means there exists $\beta_0 \in \mathcal{B}$ such that

$$H_d((-\infty, t] \,|X) = H_{d,\beta_0}((-\infty, t] \,|X^\top \beta_0), \text{ almost surely (a.s.) } \forall d \in \{0, 1\},\ t \in \mathbb{R}.$$

In general, the parameter value of β_0 is unknown and has to be estimated. For this purpose, in section IV.1, we propose a new estimation method to derive a $\sqrt{n}-$asymptotically normal semiparametric estimator of β_0. The covariates could be unbounded and no trimming is required. For the moment, let us consider that β_0 is given.

III.2 Conditional single-index Kaplan-Meier functionals

If C is a random right-censoring variable and $Y = T \wedge C,\ \delta = \mathbf{1}\{T \leq C\}$, one can link the conditional laws of C and T to the conditional subdistributions of the observations. More precisely, for any $x \in \mathcal{X}$ and $t \in \mathbb{R}$, we have

$$\begin{aligned} H_1((-\infty, t] \mid x) &= \int_{(-\infty,t]} \mathbb{P}(C \geq T \mid X = x, T = s) F_T(ds \mid x), \\ H_0((-\infty, t] \mid x) &= \int_{(-\infty,t]} \mathbb{P}(T > C \mid X = x, C = s) F_C(ds \mid x), \end{aligned} \qquad \text{(III.2)}$$

where $F_T(dt \mid x)$ and $F_C(dt \mid x)$ are the conditional distributions functions of T and C given that $X = x$. To be able to build consistent estimates of the quantities of

interest from the data, some identification assumptions are required for any value x. For this purpose one could use the usual conditional independence assumptions imposed in survival analysis

$$T \perp C \mid X. \tag{III.3}$$

For the sake of simplicity, in the following we also assume $\mathbb{P}(C = T) = 0$. Then the equations (III.2) become

$$\begin{aligned} H_1(dt \mid x) &= F_C([t, \infty) \mid x) F_T(dt \mid x) \\ H_0(dt \mid x) &= F_T([t, \infty] \mid x) F_C(dt \mid x). \end{aligned}$$

Following our dimension reduction approach, we assume

$$(T, C) \perp X \mid X^\top \beta_0 \tag{III.4}$$

for some vector $\beta_0 \in \mathcal{B}$. Our model assumptions imply that β_0 is also the vector satisfying the condition (III.1). Moreover, the equations (III.2) could be rewritten under the form

$$\begin{aligned} H_{1,\beta}(dt \mid z) &= F_{C,\beta}([t, \infty) \mid z) F_{T,\beta}(dt \mid z) \\ H_{0,\beta}(dt \mid z) &= F_{T,\beta}([t, \infty] \mid z) F_{C,\beta}(dt \mid z) \end{aligned} \tag{III.5}$$

with $\beta = \beta_0$, where for any $t \in \mathbb{R}$ and $z \in \mathbb{R}$,

$$F_{T,\beta}((-\infty, t] \mid z) = \mathbb{P}(T \leq t \mid X^\top \beta = z), \quad F_{C,\beta}((-\infty, t] \mid z) = \mathbb{P}(C \leq t \mid X^\top \beta = z),$$

and $F_{T,\beta}(dt \mid z)$ and $F_{C,\beta}(dt \mid z)$ are the associated measures.

It is well known that, for any fixed β, the equations (III.5) could be explicitly solved for $F_{T,\beta}$ and $F_{C,\beta}$. Indeed, let us consider the conditional cumulative hazard measures

$$\Lambda_{T,\beta}(dt \mid z) = \frac{F_{T,\beta}(dt \mid z)}{F_{T,\beta}([t, \infty] \mid z)} \quad \text{and} \quad \Lambda_{C,\beta}(dt \mid z) = \frac{F_{C,\beta}(dt \mid z)}{F_{C,\beta}([t, \infty) \mid z)}, \qquad z \in \mathbb{R}.$$

Then the model equations (III.5) could be solved for any z and t such that $H_\beta([t,\infty) \mid z) > 0$ and this yields

$$\Lambda_{T,\beta}(dt \mid z) = \frac{H_{1,\beta}(dt \mid z)}{H_\beta([t,\infty) \mid z)} \quad \text{and} \quad \Lambda_{C,\beta}(dt \mid z) = \frac{H_{0,\beta}(dt \mid z)}{H_\beta([t,\infty) \mid z)},$$

where $H_\beta(\cdot \mid z) = H_{0,\beta}(\cdot \mid z) + H_{1,\beta}(\cdot \mid z)$. Then, we could define the *single-index Kaplan-Meier functionals*

$$\begin{aligned} F_{T,\beta}((t,\infty] \mid z) &= \prod_{-\infty<s\leq t} \{1 - \Lambda_{T,\beta}(ds \mid z)\}, \\ F_{C,\beta}((t,\infty) \mid z) &= \prod_{-\infty<s\leq t} \{1 - \Lambda_{C,\beta}(ds \mid z)\}, \qquad t \in \mathbb{R}, \end{aligned} \tag{III.6}$$

where $\prod_{s\in A}$ stands for the product-integral over the set A (see Gill & Johansen, 1990). Let us remember that although Y takes finite values on the real line, we allow for a positive probability for the event $\{T = \infty\}$. Hence, implicitly we also assume that the conditional subdistributions H_0 and H_1 are such that $F_{C,\beta}((-\infty,\infty) \mid z) = 1$, for any z. This mild condition is satisfied only if, for each value of the covariate, the upper bound of support of $H_1(\cdot \mid x)$, that could be finite or infinite at this stage, is smaller or equal to the upper bound of the support of $H_0(\cdot \mid x)$.

Let us point out that, for each fixed vector β, *any* two conditional sub-probabilities H_0 and H_1 satisfying the mild condition $F_{C,\beta}((-\infty,\infty) \mid z) = 1$, for any z, define uniquely the functionals $F_{T,\beta}(\cdot \mid z)$ $F_{C,\beta}(\cdot \mid z)$. If the assumptions (III.4) and (III.3) hold true, then $F_{T,\beta_0}(\cdot \mid z)$ is precisely the single-index conditional probability distribution of T. In particular, this remark explains why in the following asymptotic results it will not be needed to impose the assumptions (III.4) and (III.3) and only the single-index condition (III.1) will be required.

Semiparametric estimators for $\Lambda_{T,\beta_0}(\cdot \mid x^\top\beta_0)$ and $F_{T,\beta_0}(\cdot \mid x^\top\beta_0)$ are easily built by plugging-into the previous formulae an estimator of β_0 and a nonparametric estimator of the conditional subdistributions $H_{d,\beta}(\cdot \mid x^\top\beta)$, $d \in \{0,1\}$. This idea will be detailed in the following sections. For now, let us cast new light on some recent single-index approaches in survival analysis literature.

III.3 Comparison with the previous approaches

Conditional distribution single-index models under random right censoring were recently proposed by Bouaziz & Lopez (2010) and Strzalkowska-Kominiak & Cao (2013). See also Lu (2010) and Strzalkowska-Kominiak & Cao (2014). These contributions impose the existence of a conditional density for Y and use a likelihood criterion to estimate the index β_0. Such likelihood approaches require quite involved technical assumptions. Our approach is simpler and could provide interesting insight in their identification assumptions, as is shown in the following.

A possible way to solve the system of equations (III.2) for F_T is to follow Stute (1993) and suppose that

$$T \perp C \qquad \text{and} \qquad \mathbb{P}(C \geq T \mid X, T) = \mathbb{P}(C \geq T \mid T). \tag{III.7}$$

Then the equation (III.2) becomes

$$H_1(dt \mid x) = F_C([t, \infty)) F_T(dt \mid x),$$

where $F_C(\cdot)$ is the marginal distribution function of C. Basically, the likelihood of Bouaziz & Lopez (2010) involves only this equation. To implement their estimation method, Bouaziz & Lopez (2010) replaced $F_C(\cdot)$ by the classical Kaplan-Meier estimator. Let us point out that a single-index assumption on $F_T(dt \mid x)$ induces a single-index structure on $H_1(dt \mid x)$, but also on $H_0(dt \mid x)$.

Strzalkowska-Kominiak & Cao (2013) also imposed the conditions (III.7) but they aimed at using a more adapted likelihood which corresponds to involving both equations (III.2) in the construction of the likelihood. Therefore they also imposed the condition $T \perp C \mid X$. Then one has the relationships

$$H_1(dt \mid x) = F_C([t, \infty)) F_T(dt \mid x), \qquad H_0(dt \mid x) = F_T([t, \infty) \mid x) F_C(dt). \tag{III.8}$$

Next, the single-index condition is imposed on the density of $F_T(dt \mid x)$ and the index estimated by maximum likelihood. Again, for implementation $F_C(\cdot)$ was replaced by the Kaplan-Meier estimator. Note that the assumptions of Strzalkowska-Kominiak

& Cao (2013) induce a single-index assumption on *both* $H_1(dt \mid x)$ and $H_0(dt \mid x)$.

In both approaches we discussed above, one could use the observations (Y, δ) to estimate much easily the index β_0, for instance using the estimation method we present in the next session, and then invert the equations above to recover $F_T(dt \mid x)$.

Let us also provide an interpretation of the contribution of Xia *et al.* (2010) in the case of a single-index assumption. Under the condition $T \perp C \mid X$, one has $H([t, \infty] \mid x) = F_T([t, \infty] \mid x) F_C([t, \infty] \mid x)$ so that one could rewrite the equation (III.2) under the form

$$\frac{h_1(t \mid x)dt}{H([t, \infty] \mid x)} = \frac{F_T(dt \mid x)}{F_T([t, \infty] \mid x)} = \lambda_T(t \mid x)dt,$$

where h_1 is the Radon-Nikodym derivative of H_1, that is $H_1(t \mid x) = h_1(t \mid x)dt$, and $\lambda_T(\cdot \mid x)$ is the conditional hazard function of T given $X = x$. In the case of a single-index setup, $T \perp X \mid X^\top \beta_0$ and this induces the single-index condition on the conditional hazard function, that is $\lambda_T(\cdot \mid x) = \lambda_T(\cdot \mid x^\top \beta_0)$. Next, by elementary properties of the convolution, for a kernel $\Gamma(\cdot)$ and a bandwidth b,

$$\frac{\mathbb{E}\left[\delta \Gamma((Y - t)/b)/b \mid X = x\right]}{H([t, \infty] \mid x)} \to \frac{h_1(t \mid x)dt}{H([t, \infty] \mid x)} = \lambda_T(t \mid x^\top \beta_0), \quad \text{as } b \to 0.$$

To estimate β_0, Xia *et al.* (2010) used an average-derivative approach based on the response $\delta b^{-1} \Gamma((Y - t)/b) H^{-1}([t, \infty] \mid x)$ with t running over the whole sample space of Y. In their section 3.2, the authors proposed a refined estimation by imposing the single-index condition on $H([t, \infty] \mid x)$ too; see also their assumption E2. In view of the equations (III.8) one can easily realize that, in the single-index case, our dimension reduction assumption (III.4) is practically the same as that of Xia *et al.* (2010). However, our estimation approach is simpler and could be implemented under milder assumptions. Moreover, as explained in section VII below, we could also consider multi-index assumptions and investigate other types of censoring mechanisms.

IV The index estimation method

The observations $(Y_i, \delta_i, X_i) \in \mathbb{R} \times \{0,1\} \times \mathbb{R}^p$ $1 \leq i \leq n$, are independent copies of (Y, δ, X).

IV.1 Semiparametric estimation of the index

For identification purposes, for any $\beta \in \mathcal{B}$ we fix the first component of β equal to 1. Hence $\mathcal{B} \subset \{1\} \times \mathbb{R}^{p-1}$ and we suppose that $\beta_0 \in \mathcal{B}$ is the unique vector satisfying condition (III.1). Let f_β be the density of $X^\top \beta$ that is supposed to exist. Let us point out that $X^\top \beta$ may have a density for any $\beta \in \mathcal{B}$ even if X is not an absolutely continuous vector of covariates. In particular, discrete covariates are allowed.

Next, let

$$U(t, d; \beta) = \Big\{ \mathbf{1}\{Y \leq t, \delta = d\} - H_{d,\beta}([0, t] \mid X^\top \beta) \Big\} f_\beta(X^\top \beta),$$

with $t \in \mathbb{R}$, $d \in \{0,1\}$, $\beta \in \mathcal{B}$, and let $U_i(t, d; \beta)$ be the same map applied to (Y_i, δ_i, X_i). We have the following obvious equivalence

$$\mathbb{E}[U(t, d; \beta) \mid X] = 0 \ \ a.s. \ \ \forall (t, d) \in \mathbb{R} \times \{0,1\} \quad \Leftrightarrow \quad (Y, \delta) \perp X \mid X^\top \beta. \tag{IV.1}$$

Hence β_0 is the value of the parameter that makes the zero-mean conditional expectation conditions on the left-hand side of the equivalence to hold. Then, the idea is to replace that conditional expectation conditions indexed by t and k by a more convenient marginal moment condition. Let $\omega(\cdot)$ be some function defined on $\mathbb{R}^p$ and having strictly positive integrable Fourier Transform. Define

$$I(\beta) = \int_{\mathbb{R} \times \{0,1\}} \mathbb{E}\left[\omega(X_1 - X_2) U_1(t, d; \beta) U_2(t, d; \beta)\right] d\mu(t, d), \tag{IV.2}$$

where μ is some measure on $\mathbb{R} \times \{0,1\}$. Following the idea of Li & Patilea (2014),

it can be shown that

$$I(\beta) \geq 0 \text{ and } I(\beta) = 0 \quad \text{if and only if} \quad \mathbb{E}[U(t,d;\beta) \mid X] = 0 \ \ a.s. \ \ \forall t \in \mathbb{R}, \ d \in \{0,1\}. \tag{IV.3}$$

The justification of the statement (IV.3) is provided in the Appendix. That means

$$I(\beta_0) = 0 \qquad \text{and} \qquad I(\beta) > 0, \ \ \forall \beta \in \mathcal{B}, \ \beta \neq \beta_0.$$

The estimation idea is to make some choice for $\omega(\cdot)$ and μ, to build a sample based approximation of $I(\beta)$ and to minimize it with respect to β.

Let us take $\omega(x) = \exp(-\|x\|^2/2)$, $x \in \mathbb{R}^p$, and μ equal to $F_{Y,\delta}$ the distribution function of the observations (Y,δ). Since $F_{Y,\delta}$ is unknown, in the applications one could approximate it by $\widehat{F}_{n,Y,\delta}$ the empirical distribution of the sample $(Y_1,\delta_1),\cdots,(Y_n,\delta_n)$. We will show that this substitution does not affect the asymptotic results. For $t \in \mathbb{R}$ and $d \in \{0,1\}$, let

$$\widehat{U}_i(t,d;\beta) = \frac{1}{n-1}\sum_{k=1}^{n} \left\{ \mathbf{1}\{Y_i \leq t; \delta_i = d\} - \mathbf{1}\{Y_k \leq t; \delta_k = d\} \right\} \frac{1}{g} L_{ik}(\beta, g), \tag{IV.4}$$

where $L_{ik}(\beta,g) = L((X_i - X_k)^\top \beta / g)$ and $L(\cdot)$ is a bounded univariate kernel. Let us define the empirical approximation of $I(\beta)$ as

$$\widehat{I}_n(\beta) = \frac{1}{n^2} \sum_{1 \leq i,j \leq n} \left[\frac{1}{n} \sum_{1 \leq l \leq n} \widehat{U}_i(Y_l, \delta_l; \beta) \widehat{U}_j(Y_l, \delta_l; \beta) \right] \omega_{ij},$$

where $\omega_{ij} = \omega(X_i - X_j)$. Then the estimator of β is defined as

$$\widehat{\beta} = \arg\min_{\beta \in \mathcal{B}} \widehat{I}_n(\beta). \tag{IV.5}$$

Proposition 4.1. *Suppose that the condition (III.1) and Assumption IX.1 hold true. Then $\sqrt{n}(\widehat{\beta} - \beta_0)$ converges in law to a centered multivariate normal distribution.*

The proof of Proposition 4.1 is relegated to the Appendix. Let us point out that in our proof X needs neither have a density nor a bounded support, and no trimming is required in the estimation procedure, as commonly imposed for estimating single-

index regression models. See, for instance, Horowitz (2009). Moreover, the technical conditions only concern the observed variables (Y, δ, X) and the kernel $L(\cdot)$.

The asymptotic variance of $\widehat{\beta}$ has a complicated form presented in the proof of Proposition 4.1. A convenient bootstrap procedure designed to approximate the asymptotic law of $\widehat{\beta}$ will be proposed in the following. However, for the purpose of studying the theoretical properties of the nonparametric estimates of the conditional Kaplan-Meier functionals like $F_{T,\beta_0}(\cdot \mid z)$, one only needs to know that $\widehat{\beta}$ converges β_0 at the rate $O_{\mathbb{P}}(n^{-1/2})$.

IV.2 Implementation aspects

Dimension reduction is a convenient vehicle to go beyond the parametric modeling with covariates and use more flexible approaches in an effective way. In this section we discuss some aspects related to the implementation of our dimension reduction inference approach in survival data analysis. More precisely, we complete our estimation methods with indications how the dimension reduction assumption could be tested and confidence intervals for β_0 could be built. Before proceeding to that, let us mention that one could easily remark from the proof that the conclusion of Proposition 4.1 remains valid if ω_{ij} is replaced by $\omega_{ij}\mathbf{1}\{i \neq j\}$ in the definition of $\widehat{I}_n(\beta)$. However, our empirical experience indicates that keeping the diagonal terms in the definition of $\widehat{I}_n(\beta)$ leads to more stable numerical results. Moreover, the $\sqrt{n}-$asymptotic normality is expected to remain true even if the single-index condition (III.1) does not hold. However, in such a case β_0 has to be replaced by $\overline{\beta} = \arg\min_{\beta \in \mathcal{B}} I(\beta)$. See also Li & Patilea (2014) for a similar situation in the case without censoring.

Single-index assumption check

Single-index assumptions are very common dimension reduction devices. However, one should be able to check whether such assumptions are realistic for the data at hand. Following an approach introduced by Maistre & Patilea (2014), we propose a formal test of the single-index assumption (III.1) against general alternatives.

For any $\beta \in \mathcal{B}$ let $\mathbf{A}(\beta)$ be a $p \times (p-1)-$matrix with real entries such that the $p \times p-$matrix $(\beta\ \mathbf{A}(\beta))$ is orthogonal. The orthogonality is not necessary, invertibility suffices, but orthogonality is expected to lead to better finite sample properties for the test. Let $G(\cdot)$ be an univariate kernel with positive Fourier Transform on the real line. For some bandwidth $\mathfrak{b}$ and any $1 \leq i \neq j \leq n$, let

$$G_{ij}(\beta, \mathfrak{b}) = G((X_i - X_j)^\top \beta / \mathfrak{b}) \quad \text{and} \quad \phi_{ij}(\beta) = \exp(-\|(X_i - X_j)^\top \mathbf{A}(\beta)\|^2/2).$$

Next, with $\widehat{U}_i$ and $\widehat{U}_j$ defined as in equation (IV.4), consider

$$Q_n(\beta) \;=\; \frac{1}{n\,(n-1)\,\mathfrak{b}} \sum_{1 \leq i \neq j \leq n} \left\langle \widehat{U}_i(\cdot,\cdot,\beta),\ \widehat{U}_j(\cdot,\cdot,\beta) \right\rangle_n G_{ij}(\beta, \mathfrak{b}) \phi_{ij}(\beta),$$

where for any $u(\cdot,\cdot)$ and $v(\cdot,\cdot)$ bounded functions defined on $\mathbb{R} \times \{0,1\}$,

$$\langle u(\cdot,\cdot),\ v(\cdot,\cdot) \rangle_n = \int_{\mathbb{R}\times\{0,1\}} u(t,d) v(t,d) d\widehat{F}_{n,Y,d}(t,d) = \frac{1}{n} \sum_{l=1}^{n} u(Y_l, \delta_l) v(Y_l, \delta_l).$$

The rationale for considering $Q_n(\beta)$ is that it represents a sample based approximation of

$$Q(\beta, \mathfrak{b}) = \mathbb{E}\left[\left\{ \int_{\mathbb{R}\times\{0,1\}} U_1(t,d;\beta) U_2(t,d;\beta) dF_{Y,d}(t,d) \right\} \mathfrak{b}^{-1} G((X_1 - X_2)^\top \beta / \mathfrak{b}) \phi_{12}(\beta) \right].$$

The Inverse Fourier Transform argument used to justify the equation (IV.3) implies that for any $\mathfrak{b} > 0$, $Q(\beta, \mathfrak{b}) \geq 0$ and $Q(\beta, \mathfrak{b}) = 0$ if and only if

$$\mathbb{E}[U(t,d;\beta) \mid X] = 0, \quad a.s., \quad \forall (t,d) \in \mathbb{R} \times \{0,1\}.$$

Given the equivalence from the equation (IV.1) above, under the null hypothesis (III.1) the quantity $Q_n(\beta)$ should be close to zero if β is close to β_0. If the single-index assumption (III.1) does not hold, $Q_n(\beta)$ is expected to be uniformly bounded away from zero. Let us point out that one could obtain $Q_n(\beta)$ from $\widehat{I}_n(\beta)$ after replacing ω_{ij} by $[n/(n-1)]\mathfrak{b}^{-1} G_{ij}(\beta, \mathfrak{b}) \phi_{ij}(\beta)$. The univariate smoothing induced by this modification will allow to end with a pivotal test statistics.

The variance of $Q_n(\beta)$ could be estimated by

$$\widehat{\omega}_n(\beta)^2 = \frac{2}{n^2\left(n-1\right)^2\mathfrak{b}^2}\sum_{1\leq i\neq j\leq n}\left\langle \widehat{U}_i(\cdot,\cdot,\beta),\ \widehat{U}_j(\cdot,\cdot,\beta)\right\rangle_n^2 G_{ij}^2(\beta,\mathfrak{b})\phi_{ij}^2(\beta).$$

If $\widehat{\beta}$ is the estimator defined in equation (IV.5), then the test statistic we propose is

$$T_n = \frac{Q_n(\widehat{\beta})}{\widehat{\omega}_n(\widehat{\beta})}. \tag{IV.6}$$

By minor modifications of a result in Maistre & Patilea (2014), under some regularity conditions and a suitable rate of decrease to zero for $\mathfrak{b}$, one could show that T_n converges in law to a standard normal distribution if the single-index assumption (III.1) holds true. Moreover, the test could detect directional alternatives of the form

$$\mathbb{P}(Y\leq t,\delta=d\mid X) = \mathbb{P}(Y\leq t,\delta=d\mid X^\top\beta_0) + r_n\delta(X,t,d),\ \ (t,d)\in\mathbb{R}\times\{0,1\},$$

as soon as $r_n^2 n\mathfrak{b}^{1/2}\to\infty$.

Confidence regions

One could be interested in confidence regions for the index β_0 and the values of the conditional distribution $F_{T,\beta_0}(\cdot\mid z)$. This could be derived conveniently using numerical methods.

Following Lavergne & Patilea (2013) and Li & Patilea (2014), one could build a randomly perturbed version of $I(\beta)$ as

$$\widehat{I}_n^*(\beta) = \frac{1}{n^2}\sum_{1\leq i,j\leq n}\left[\frac{1}{n}\sum_{1\leq l\leq n}\widehat{U}_i(Y_l,\delta_l;\beta)\widehat{U}_j(Y_l,\delta_l;\beta)\right]\omega_{ij}^*, \tag{IV.7}$$

where

$$\omega_{ij}^* = \omega(X_i - X_j)\zeta_i\zeta_j,$$

with $\zeta_1,\cdots,\zeta_n$ a sample of independent exponential random variables with mean equal to 1, independent of the observations. A new estimate of β_0 could be obtained

as

$$\widehat{\beta}^* = \arg\min_{\beta\in\mathcal{B}} \widehat{I}_n^*(\beta).$$

The procedure could repeated many times and confidence regions could be derived using the sample of $\widehat{\beta}^*$'s. See, for instance, Chiang & Huang (2012), Lavergne & Patilea (2013), Li & Patilea (2014) for some theoretical justification of this type of approach for building confidence regions. A method for building confidence intervals for $F_{T,\beta_0}(\cdot \mid x^\top\beta_0)$ will be mentioned in the next section.

V Semiparametric estimators of the Kaplan-Meier functionals

With at hand an estimate $\widehat{\beta}$ of the vector β_0 and nonparametric estimates of $H_{0,\beta}(\cdot \mid x^\top\beta)$ and $H_{1,\beta}(\cdot \mid x^\top\beta)$ for arbitrary β, one could use the sample-based version of the formula (III.6) to build an estimator for the Kaplan-Meier functional $F_{T,\beta_0}(\cdot \mid x^\top\beta_0)$. The result will be precisely the conditional Kaplan-Meier estimator with univariate covariate $X^\top\widehat{\beta}$.

Following Beran (1981) and Dabrowska (1989), here we will use Nadaraya-Watson estimators for the sub-distributions of the observations. That is, for any β and $z \in \mathbb{R}$, let

$$\widehat{F}_{T,\beta}((t,\infty] \mid z) = \prod_{Y_i\le t}\left(1 - \frac{w_{in}(z;\beta)}{\sum_{j=1}^n \mathbf{1}_{\{Y_j\ge Y_i\}} w_{jn}(z;\beta)}\right)^{\delta_i}, \qquad \text{for } t \le Y_{(n)},$$

where $Y_{(n)} = \max_{1\le i\le n} Y_i$ and

$$w_{in}(z;\beta) = \frac{K((X_i^\top\beta - z)h^{-1})}{\sum_{l=1}^n K((X_l^\top\beta - z)h^{-1})},$$

$K(\cdot)$ is a kernel and h is the bandwidth. If $Y_{(n)}$ is an uncensored observation then $\widehat{F}_{T,\beta}((Y_{(n)},\infty] \mid z) = 0$, but when $Y_{(n)}$ is a censored observation, then

$$\widehat{F}_{T,\beta}([Y_{(n)},\infty] \mid z) = \widehat{F}_{T,\beta}((Y_{(n)},\infty] \mid z) > 0.$$

Under the single-index assumption (III.4), the conditional distribution function $F_T(\cdot \mid x)$ of the lifetime T given $X = x$ is equal to $F_{T,\beta_0}(\cdot \mid x^\top \beta_0)$. Hence, a natural estimator of $F_T(\cdot \mid x)$ is $F_{T,\widehat{\beta}}(\cdot \mid x^\top \widehat{\beta})$. Below, we derive a i.i.d. representation for this estimator.

V.1 Asymptotic results for Kaplan-Meier functionals

Let

$$\tau_H(x;\beta_0) = \inf\{t : H((t,\infty) \mid x^\top \beta_0) = 0\},$$

and consider $\overline{\mathcal{X}} \subset \mathcal{X}$ a compact subset in the space of the covariates such that

$$\inf_{x \in \overline{\mathcal{X}}} f_{\beta_0}(x^\top \beta_0) > 0.$$

Let $\tau > 0$ such that

$$\tau < \inf_{x \in \overline{\mathcal{X}}} \tau_H(x;\beta_0).$$

The following i.i.d. asymptotic representation is en extension of the results of Du & Akritas (2002) and Lopez (2011).

Proposition 5.1. *Assume that the single-index condition (III.1) holds true. For each $x \in \overline{\mathcal{X}}$ and $t \leq \tau$, let*

$$\eta_{\Lambda_T,i}(t, x^\top \beta_0) = w_{in}(x^\top \beta_0, \beta_0)\left(\frac{\delta_i \mathbf{1}\{Y_i \leq t\}}{H_{\beta_0}([Y_i,\infty) \mid x^\top \beta_0)} - \int_{(-\infty,t]} \frac{\mathbf{1}\{Y_i \geq s\} H_{1,\beta_0}(ds \,|\, x^\top \beta_0)}{H^2_{\beta_0}([Y_i,\infty) \mid x^\top \beta_0)} \right)$$

and

$$\eta_{F_T,i}(t, x^\top \beta_0) = -F_{T,\beta_0}((t,\infty] | \, x^\top \beta_0) w_{in}(x^\top \beta_0, \beta_0) \Bigg[\frac{\delta_i F_{T,\beta_0}([Y_i,\infty] \,|x^\top \beta_0) \mathbf{1}\{Y_i \leq t\}}{F_{T,\beta_0}((Y_i,\infty] | x^\top \beta_0) H_{\beta_0}([Y_i,\infty) | x^\top \beta_0)} - \int_{(-\infty,t]} \frac{F_{T,\beta_0}([Y_i,\infty] \mid x^\top \beta_0) \mathbf{1}\{Y_i \geq s\} H_{1,\beta_0}(ds \mid x^\top \beta_0)}{F_{T,\beta_0}((Y_i,\infty] \mid x^\top \beta_0) H^2_{\beta_0}([Y_i,\infty) \mid x^\top \beta_0)} \Bigg]$$

Let the kernel $K(\cdot)$ be a symmetric probability density function with compact support and twice continuously differentiable. Assume that for any $b_n \to 0$ and $d = 0$ or

$d = 1$,

$$\sup_{0<\|\beta-\beta_0\|\leq b_n} \left\{ \frac{|f_\beta(z) - f_{\beta_0}(z)|}{\|\beta - \beta_0\|} + \frac{|H_{d,\beta}((-\infty, t] \,|\, z) f_\beta(z) - H_{d,\beta_0}((-\infty, t] \,|\, z) f_{\beta_0}(z)|}{\|\beta - \beta_0\|} \right\} \leq C$$

where C is a constant independent of $z, t \in \mathbb{R}$. Let $\widehat{\beta}$ be a consistent estimator of β_0. Under the Assumption IX.1,

$$\begin{aligned} \widehat{\Lambda}_{T,\widehat{\beta}}((-\infty, t] \mid x^\top \widehat{\beta}) - \Lambda_{T,\beta_0}((-\infty, t] \mid x^\top \beta_0) &= \frac{1}{n} \sum_{i=1}^{n} \eta_{\Lambda_T,i}(t, x^\top \beta_0) + R_{n,\Lambda_T}(t, x) \\ \widehat{F}_{T,\widehat{\beta}}((t, \infty] \mid x^\top \widehat{\beta}) - F_{T,\beta_0}((t, \infty] \mid x^\top \beta_0) &= \frac{1}{n} \sum_{i=1}^{n} \eta_{F_T,i}(t, x^\top \beta_0) + R_{n,F_T}(t, x), \quad \text{(V.8)} \end{aligned}$$

with

$$\sup_{t \leq \tau,\, x \in \overline{\mathcal{X}}} \{|R_{n,\Lambda_T}(t, x)| + |R_{n,F_T}(t, x)|\} = O_{\mathbb{P}}(n^{-1} h^{-1} \ln n + \|\widehat{\beta} - \beta_0\|).$$

If $\widehat{\beta}$ is a $\sqrt{n}-$estimator, the reminders R_{n,Λ_T}, R_{n,F_T} have the uniform rate $O_{\mathbb{P}}(n^{-1/2})$, that is negligible compared to $O_{\mathbb{P}}(n^{-1/2} h^{-1/2})$, the rate of the i.i.d. sums.

The i.i.d. asymptotic representation is a powerful result that serves to develop asymptotic theory in various situations. See, for instance, Du & Akritas (2002) and Lopez (2011) for some examples and references. Here, we illustrate a direct consequence for a single-index cure rate estimator that extends the result of Xu & Peng (2014). Cure models received a lot of attention lately, see, for instance, Tsodikov *et al.* (2003) and Zheng *et al.* (2006). The conditional cure rate $\pi(x) = \mathbb{P}(T = \infty \mid x)$ could be estimated by the conditional Kaplan-Meier estimator of Beran (1981) taken at the largest uncensored observation. In the case of a univariate fixed-design covariate, Xu & Peng (2014) assume that

$$-\infty < \sup_x \tau_{H_1}(x) < \inf_x \tau_H(x) < \infty,$$

where

$$\tau_{H_1}(x) = \inf\{t \mid H_1((t, \infty) \mid x) = 0\}, \quad \tau_H(x) = \inf\{t \mid H((t, \infty) \mid x) = 0\},$$

so that for any x,

$$\int_{[0,\tau_{H_1}(x)]} \frac{H_1(dt \mid x)}{H^2([t,\tau_H(x)] \mid x)} < \infty.$$

Under these conditions and some technical assumptions, Xu and Peng proved that, for any $t \in [0, \tau_H(x)]$,

$$\sqrt{nh}\left(\widehat{\Lambda}_T([0,t] \mid x) - \Lambda_T([0,t] \mid x)\right) \rightsquigarrow N(0, \sigma^2(t,x)),$$

where

$$\sigma^2(t,x) = \int_{[0,t]} \frac{\Lambda_T(ds \mid x)}{H([s,\tau_H(x)] \mid x)} \int K^2(t)dt$$

and $\rightsquigarrow$ denotes convergence in law. Moreover,

$$\sqrt{nh}\left(\widehat{F}_T((Y^1_{(n)}, \infty]) \mid x) - \pi(x)\right) \rightsquigarrow N(0, \pi^2(x)\sigma^2(x)),$$

where $\widehat{F}_T$ is the Beran estimator, $Y^1_{(n)}$ is the largest uncensored observation of Y and

$$\sigma^2(x) = \int_{[0,\tau_{H_1}(x)]} \frac{\Lambda_T(ds \mid x)}{H([s,\tau_H(x)] \mid x)} \int K^2(t)dt.$$

Our i.i.d. representation allows to recover this result and to extend it to multivariate covariates using the single-index framework, as stated in the following corollary.

Corollary 5.2. *Assume that the assumptions of Proposition 5.1 and the assumptions (III.3) and (III.4) hold true. Then*

$$\pi(x) = \mathbb{P}(T = \infty \mid x) = F_{T,\beta_0}((\tau_H(x), \infty] \mid x^\top \beta_0), \qquad \forall x \in \mathcal{X}.$$

Assume

$$-\infty < \sup_{x \in \mathcal{X}} \tau_{H_1}(x; \beta_0) < \inf_{x \in \overline{\mathcal{X}}} \tau_H(x; \beta_0) < \infty,$$

where $\tau_{H_1}(x;\beta_0) = \inf\{t \mid H_{1,\beta_0}((t,\infty) \mid x^\top\beta_0) = 0\}$. *If* $x \in \overline{\mathcal{X}}$ *and* $\widehat{\beta}$ *is a* $\sqrt{n}-$*consistent estimator of* β_0,

$$\sqrt{nh}\left(\widehat{F}_{T,\widehat{\beta}}((Y^1_{(n)}, \infty]) \mid x^\top\widehat{\beta}) - \pi(x)\right) \rightsquigarrow N(0, \pi^2(x)\sigma^2_{\beta_0}(x^\top\beta_0))$$

$$\sigma^2_{\beta_0}(x) = \int_{(-\infty,\tau_{H_1}(x;\beta_0)]} \frac{H_{1,\beta_0}(ds \mid x^\top\beta_0)}{H^2_{\beta_0}([s,\tau_H(x)] \mid x^\top\beta_0)} \int K^2(t)dt.$$

To build confidence regions for $F_{T,\beta_0}((-\infty,t] \mid x^\top\beta_0)$ for given x and t, one could resample (Y,δ) from the subdistributions $H_{1,\widehat{\beta}}$ and $H_{0,\widehat{\beta}}$. This could be done under a fixed design assumption, as van Keilegom & Veraverbeke (1997) did in the case of a univariate covariate, or taking into account a univariate random design as in Li & Datta (2001). We argue that given the parametric convergence rate of $\widehat{\beta}$, the validity of such bootstrap procedures could be derived by similar arguments as used in the univariate case. The detailed investigation of this issue will be considered elsewhere.

VI Empirical evidence

To illustrate the new approach, we performed and present in the following few empirical experiments with simulated and real data.

VI.1 Simulation experiments

We consider a lifetime T defined as $T = 5(X^\top\beta_0)^2 + \varepsilon$ with a three-dimensional vector of covariates $X = (X^{(1)}, X^{(2)}, X^{(3)})^\top$, where $(X^{(1)}, X^{(2)})^\top$ are generated from a zero mean bivariate normal with standard deviation equal to 1 and correlation equal to 0.2, and $X^{(3)}$ is a Bernoulli random variable with parameter $p = 0.4$. The variable ε is standard normal and independent of X. The true value of the parameter is $\beta_0 = (\beta_{01}, \beta_{02}, \beta_{03})^\top = (1,1,1)^\top$. The censoring variable is $C = c\,\eta\,\exp(\sqrt{|X^\top\beta_0|/5}) - 3$ with η uniformly distributed on $[0,1]$ and c a constant which controls the proportion of censoring. In our simulation, we use $c = 5.8$ and 20.3 which yields approximately 80% and 40% censoring proportion, respectively. Let us recall that in order to guarantee identification, we set the first component of $\beta = (\beta_1, \beta_2, \beta_3)^\top$ equal to 1.

First, we proceed to the estimation of the index parameter β_0 using the estimator $\widehat{\beta} = (1, \widehat{\beta}_2, \widehat{\beta}_3)^\top$ proposed in equation (IV.5). We compare our method to the

hMAVE (MAVE for hazard function) proposed by Xia *et al.* (2010). They prefer to identify the parameter β by constraining the norm to be equal to 1 (and fixing the sign of one component). Here, in order to have comparable results, we divide the estimate of β obtained with their method by its first component. (We also performed simulations, not reported here, using Xia *et al.* (2010)'s identification condition in our equation (IV.5). The conclusions are very similar.) To compare performances, we report several statistics obtained with 500 independent samples of size $n = 50$: the mean, the standard deviation, the median and the mean squared error of the estimators of β_2 and β_3, as well as the mean and the standard deviation of the largest singular value (sv) of $\widehat{\beta}\widehat{\beta}^\top - \beta_0\beta_0^\top$ and of the absolute estimation errors (aee) $|\widehat{\beta_2} - \beta_{02}| + |\widehat{\beta_3} - \beta_{03}|$. The kernel $L(\cdot)$, which is used to define $\widehat{U}_i(t, d; \beta)$ in equation (IV.4), is the gaussian kernel. The bandwidth for this kernel is selected as the minimizer of the loss $\widehat{I}_n(\widehat{\beta})$ with respect to g on an equidistant grid $\{0.01, 0.03, \cdots, 0.13\}$. The simulation results are shown in Table 4.1. Our method performs quite well. Compared with hMAVE, it has a similar absolute bias, but the law of our estimator seems more concentrated.

Table 4.1: Descriptive statistics for the estimators of $\beta_0 = (1, 1, 1)^\top$, and the mean and the standard deviation (in parentheses) of the *aee* and *sv* values, obtained with 500 independent samples of size $n = 50$. The results obtained with the method of Xia *et al.* (2010) are presented in the gray cells.

	80% censoring							
	mean		std		median		mse	
	ours	hMAVE	ours	hMAVE	ours	hMAVE	ours	hMAVE
β_2	1.0539	0.9838	0.2902	0.4214	0.9876	0.9920	0.0869	0.1775
β_3	0.9819	0.9435	0.2383	0.5029	0.9724	0.9678	0.0570	0.2556
aee	0.2985(0.4000)				0.6260(0.5430)			
sv	0.6730(1.1801)				1.3661(1.3603)			
	40% censoring							
	mean		std		median		mse	
	ours	hMAVE	ours	hMAVE	ours	hMAVE	ours	hMAVE
β_2	0.9955	1.0011	0.1934	0.2249	0.9691	0.9816	0.0373	0.0505
β_3	0.9813	0.9844	0.1608	0.2890	0.9735	0.9737	0.0261	0.0836
aee	0.2133(0.2529)				0.3960(0.2547)			
sv	0.4584(0.7236)				0.8828(0.6637)			

Next, we used the idea described in section IV.2 above to build confidence in-

tervals for β_{02} and β_{03}. For each of the 500 samples of size $n = 50$ we generated 299 independent random samples $\zeta_1, \cdots, \zeta_n$ and computed the criteria $\widehat{I}_n^*(\beta)$ as in equation (IV.7). Only the setup with 80% of censoring is reported. The 90% and 95% confidence intervals obtained with the optimal values $\widehat{\beta}_2^*$ and $\widehat{\beta}_3^*$ are presented in Table 4.2. The levels for β_{02} are slightly less than the nominal ones, while the intervals for the coefficient of the discrete component of X are slightly larger than necessary. Overall, the approach we propose for building confidence intervals performs reasonably well with small samples where the asymptotic approximation could be quite poor.

Table 4.2: Small sample confidence intervals: a number of 299 perturbed criteria $\widehat{I}_n^*(\beta)$ as defined in equation (IV.7) are used for any of the 500 samples of size $n = 50$. The censoring amount is 80%.

	90% confidence interval		95% confidence interval	
	length	percentage	length	percentage
β_2	0.9341	87.2	1.1610	95.4
β_3	1.0859	93.8	1.3315	97.6

Finally, we considered the estimation of $F_{T,\beta_0}((t,\infty] \mid x) = F_{T,\beta_0}((t,\infty] \mid x^\top\beta_0)$ using the conditional Kaplan-Meier estimator. The values of t considered are the quantiles 0.1, 0.2,...,0.9 of the true conditional law. To approximate β_0 we considered both cases, our estimator and Xia *et al.* (2010)'s estimator. We fixed two vectors x, namely $x = (-0.5, -0.5, 1)^\top$ and $x = (0.1, -0.3, 0)^\top$, for which we considered the respective amount of censoring of 80% and 40%. The quartic (biweight) kernel $K(z) = (15/16)(1 - z^2)^2\mathbf{1}\{|z| \leq 1\}$ is used, and the bandwidth h is taken equal to 0.4 with each of the two estimates of β_0 we use. (The quartic kernel is not twice differentiable as required in our Proposition 5.1, but since our empirical experience shows that this incoherence has practically no impact, we keep it for reasons of comparison with previous contributions in the literature.) The results are presented in Tables 4.3 and 4.4, those obtained with Xia *et al.*'s estimate of β_0 are again in gray cells. The estimates based on our $\widehat{\beta}$ are less biased and overall quite accurate.

Table 4.3: Single-index estimator of the conditional distribution: 500 samples of size $n = 50$, censored rate is 80%, $x = (-0.5, -0.5, 1)^\top$

$F_{T,\beta_0}((t,\infty] \mid x^\top\beta_0)$		0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
$\widehat{F}_{T,\widehat{\beta}}((t,\infty] \mid x^\top\widehat{\beta})$	ours	0.921	0.838	0.757	0.670	0.587	0.497	0.400	0.301	0.180
	hMAVE	0.943	0.883	0.819	0.742	0.669	0.594	0.506	0.410	0.295

Table 4.4: Single-index estimator of the conditional distribution: 500 samples of size $n = 50$, censored rate is 40%, $x = (0.1, -0.3, 0)^\top$

$F_{T,\beta_0}((t,\infty]\|x^\top\beta_0)$		0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
$\widehat{F}_{T,\widehat{\beta}}((t,\infty]\|x^\top\widehat{\beta})$	ours	0.906	0.819	0.729	0.631	0.544	0.455	0.361	0.261	0.153
	hMAVE	0.906	0.830	0.745	0.650	0.573	0.491	0.398	0.298	0.187

VI.2 Real data application

For a real data illustration we consider the well-known Stanford heart transplantation data set given by Miller & Halpern (1982). The data set includes survival times and covariates for 184 patients who had received a heart transplantation between October 1967 and February 1980. For reasons of comparison with previous empirical work, see for instance Miller and Halpern (1982), Wei *et al.* (1990) and van Keilegom *et al.* 2001), we restrict our attention to the 157 out of 184 individuals who had complete tissue typing. Moreover, the response variable will be taken equal to the base 10 logarithm of the survival time (in days). The covariates will be the age of the patient at the time of the first heart transplant and the T5 mismatch score which measure the degree of tissue incomputability between the initial donor and recipient hearts with respect to HLA antigens. A set of 55 patients alive beyond February 1980 were considered censored.

For identification with our single-index approach, the coefficient of the covariate 'age' is fixed equal to 1. The estimate of the coefficient of the covariate 'T5 mismatch score' is equal to 2.9333. This value may seem surprising given the previous analysis in the literature, but one has to keep in mind the normalization induced by the identification assumption. The 95% confidence interval is (0.2972, 3.0769) which suggests that this covariate could be significant. The effect of the T5 mismatch score is not captured by the classical models which indicate that the coefficient of this covariate is not significant (see Table 4.5, the comparison of previous methods.) Next, we applied the conditional Kaplan-Meier estimator with the index $(1, 2.9333)^\top$

and several values of the covariates, using the quartic kernel. The estimated curves are presented in the four panels of Figure 4.1. In panel (a) we considered the bandwidth $h = 7$, as considered by van Keilegom *et al.* (2001), and we set the covariate 'age' equal to its empirical mean. The curve corresponding to a null value for the covariate 'T5 mismatch score', as well as the curves corresponding to three non-null values of the covariate 'T5 mismatch score' are presented. The differences between the four curves advocate for an effect of the T5 mismatch score. In panel (b) we investigate the effect of a change in the bandwidth h when 'age' is equal to 30 and 'T5 mismatch score' is equal to its empirical mean. Quite little change could be noticed when h varies from 6.5 to 7 and to 7.5. Once again, the effect of 'T5 mismatch score' seems significant in view of the curve corresponding to the absence of this covariate, also plotted in the panel (b). The experience from panel (b) was repeated with 'age' equal to 40 and 50 and the results are presented in the panels (c) and (d), respectively. The same conclusion could be drawn: the effect of a slight change in the bandwidth h seems small. Meanwhile, the effect of the covariate 'T5 mismatch score' could be more important, suggesting that this covariate is significant.

Table 4.5: The estimats and 95% confidence interval of the covariates in Stanford heart transplantation data set

	Age		T5	
	estimate	confidence interval	estimate	confidence interval
Ours	1	—	2.933	(0.297, 3.077)
Wei *et al.* (1990)	-0.025 -0.021	(-0.047,-0.007) (-0.041,-0.004)	-0.124 -0.062	(-0.395,0.197) (-0.331,0.248)
Buckley & James (1979)	-0.015	(-0.031,0.001)	-0.003	(-0.266,0.260)
Miller & Halpern (1982)	0.000	(-0.016,0.016)	0.040	(-0.225,0.305)
Cox	0.030	(0.008,0.052)	0.167	(-0.192,0.526)

VII Discussion and extensions

We propose a new way to build estimators under dimension reduction assumptions in survival analysis with covariates. The main idea is to impose the dimension reduction assumption in the observations space and to estimate the conditional law

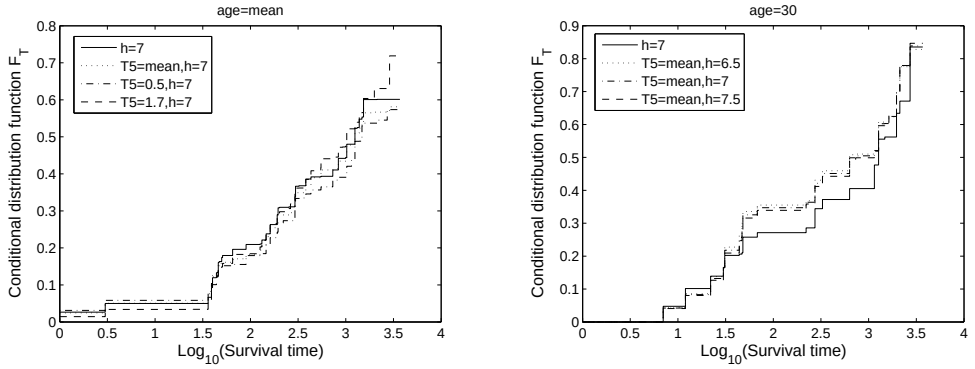


Figure 4.1: The estimates of the conditional distribution function. The continuous lines correspond to the estimates obtained using with only the covariate 'age', as considered by van Keilegom *et al.* (2001).

of the observations under such conditions. Here, we consider a well-known dimension reduction approach based on indices, that is we assume that all the information on the responses carried by the covariates is contained in a small number of linear combinations of them. These linear combinations of the covariates is given by unknown index vectors. It is worthwhile to note that such a dimension reduction introduced through the observations space could be more easily tested, for instance following the approach introduced by Maistre & Patilea (2014). Next, the idea is to use the map generated by the censoring mechanism that links the conditional law of the observations to the quantities of interest and to plug-in the estimates of the conditional law of the observations. As a result, we obtain easy to calculate semiparametric estimates of the conditional law of the lifetime of interest and of any other smooth functional of the observations. Moreover, bootstrapping in the space of the observations and applying the map to the conditional law estimate obtained with each bootstrap sample, we could easily build confidence regions for the quantities of interest. In this paper, we focus on single-index assumptions and random right censoring but our principle is much more general, as we explain below. The asymptotic properties of the estimators could be rapidly derived and under mild conditions. We also showed that the related contributions in the literature eventually lead to a single-index structure on the conditional law of the observations, like the one we consider herein. However, the existing estimation procedures are much more complicated due to the fact that they involve simultaneously the two aspects: dimension reduction and the map generated by the censoring mechanism linking the observations to the quantities of interest.

Let us now suggest some possible extensions. Other asymptotic results already proved for the conditional Kaplan-Meier approach could be adapted to the dimension reduction idea to derive single-index versions of them. For instance, one could consider a single-index version of the Bahadur-type representation derived by Dabrowska (1992) for the estimator of the conditional quantile of T obtained by inversion of the classical conditional Kaplan-Meier estimator.

Clearly, our dimension reduction approach could be extended to multi-index

assumptions. Suppose there exists a $p \times r-$matrix B_0 with $1 \leq r < p$ such that

$$(Y, \delta) \perp X \mid X^\top B_0. \tag{VII.1}$$

A similar condition is used by Xia *et al.* (2010). To estimate a basis in the space generated by the columns of B_0, it suffices to redefine the $\widehat{U}_i$ introduced in equation (IV.4) as

$$\widehat{U}_i(t, k; B) = \frac{1}{n-1} \sum_{k=1}^{n} \left\{ \mathbf{1}\{Y_i \leq t; \delta = d\} - \mathbf{1}\{Y_k \leq t; \delta = d\} \right\} \frac{1}{g^r} L_{ik}(B, g),$$

with $L_{ik}(B, g) = L((X_i - X_k)^\top B/g)$ and $L(\cdot)$ a $r-$dimension kernel. Next, define $I_n(B)$ accordingly and minimize $I_n(B)$ with respect cu B under suitable identification restriction (for instance, the top block of B could be fixed equal to the $r \times r$ identity matrix). As pointed out by Maistre & Patilea (2014), the test statistic T_n defined in equation (IV.6) could be modified in order to test the null hypothesis (VII.1) of a $r-$dimension multi-index hypothesis. With at hand an estimate $\widehat{B}$, one can build the conditional Kaplan-Meier estimator with $r-$dimension covariate vectors $X_i^\top \widehat{B}$, just like in Beran (1981) and Dabrowska (1989).

Finally, let us point out that in principle our approach could apply to a larger class of censoring mechanisms. The implementation and the theoretical investigation of the estimators is much easier when the map linking the conditional law of the observations to the quantities of interest has a closed form expression. Let us end with another example of censoring mechanism considered by Patilea & Rolin (2006) that generates a closed form expression map between the observations and a lifetime of interest $T \in [0, \infty]$. The observations are independent copies of (Y, A, X) where X is a vector of p covariates, $A \in \{0, 1, 2\}$, $Y \in [0, \infty)$ and

$$\begin{cases} Y < T & \text{if} \quad A = 0, \\ Y = T & \text{if} \quad A = 1, \\ Y > T & \text{if} \quad A = 2. \end{cases}$$

Such observations could occur for instance when the age of the occurrence of a disease

is under study and only one of the following situation is possible: a) evidence of the disease is present and the age at onset is known (from medical records, interviews with the patient or family members,...); b) the disease is diagnosed but the age at onset is unknown or the accuracy of the information about this is questionable; and c) the disease is not diagnosed at the examination time. Following our approach, assume that $(Y, A) \perp X \mid X^\top \beta_0$ for some $\beta_0 \in \mathbb{R}^p$. For $\beta \in \mathbb{R}^p$, $0 \leq t < \infty$, define the sub-distributions

$$H_{k,\beta}([0,t] \mid z) = P(Y \leq t, A = k \mid X^\top \beta = z), \quad k = 0, 1, 2, \;\; z \in \mathbb{R}.$$

The conditional law of the observations is given by $H_{k,\beta_0}(\cdot \mid z)$, $k = 0, 1, 2$. Let C be a censoring time and $\Delta \in \{0, 1\}$ a 'latent' variables that models the fact that T is observed or not when $T \leq C$. That means

$$Y = \min(T, C) + (1 - \Delta) \max(C - T, 0) = C + \Delta \min(T - C, 0)$$

and $A = 2(1 - \Delta)\mathbf{1}\{T \leq C\} + \mathbf{1}\{C < T\}$. Let $p_\beta(z) = \mathbb{P}(\Delta = 1 \mid X^\top \beta = z)$, $z \in \mathbb{R}$. If T, C and Δ are independent given X, for any z we can write

$$\begin{cases} H_{0,\beta_0}(dt \mid z) &= F_{T,\beta_0}((t,\infty] \mid z) F_{C,\beta_0}(dt \mid z), \\ H_{1,\beta_0}(dt \mid z) &= p_{\beta_0}(z) F_{C,\beta_0}([t,\infty] \mid z) F_{T,\beta_0}(dt \mid z), \\ H_{2,\beta_0}(dt \mid z) &= (1 - p_{\beta_0}(z)) F_{T,\beta_0}([0,t] \mid z) F_{C,\beta_0}(dt \mid z). \end{cases}$$

The case $p \equiv 1$ corresponds to the usual right-censoring setup, while $p \equiv 0$ corresponds to current-status data. If $p_{\beta_0}(z) > 0$ the system could be solved explicitly to obtain

$$p_{\beta_0}(z) = \frac{H_{0,\beta_0}([0,\infty) \mid z)}{H_{0,\beta_0}([0,\infty) \mid z) + H_{2,\beta_0}([0,\infty) \mid z)}$$

and

$$\Lambda_{T,\beta_0}(dt \mid z) = \frac{H_{0,\beta_0}(dt \mid z)}{H_{0,\beta_0}([t,\infty) \mid z) + p_{\beta_0}(z) H_{1,\beta_0}([t,\infty) \mid z)}.$$

Moreover, one could allow for a positive probability of the event $\{T = \infty\}$ and write

$$\mathbb{P}(T = \infty \mid X^\top \beta_0 = z) = \prod_{t \in (0,\infty)} \{1 - \Lambda_{T,\beta_0}(dt \mid z)\}.$$

An estimate $\widehat{\beta}$ for β_0 could be build by an obvious extension of the method proposed in section IV.1. Then one can easily estimate $H_{k,\widehat{\beta}}([0,t] \mid x^\top \widehat{\beta})$, $k = 0, 1, 2$, by kernel smoothing and plug-in the estimates in the formulae above to obtain estimates for the quantities of interest.

VIII Reference

1. Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley.
2. Bouaziz, O. & Lopez, O. (2010). Conditional density estimation in a censored single-index model. *Bernoulli* **16**, 514–542.
3. Buckley, J. & James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429–436.
4. Chiang, C.-T. & Huang, M.-Y. (2012). New estimation and inference procedures for a single-index conditional distribution model. *J. Multivariate Anal.* **111**, 271–285.
5. Cox, D.R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187–220.
6. Dabrowska, D. M. (1989). Uniform Consistency of the Kernel Conditional Kaplan-Meier Estimate. *Ann. Statist.* **17**, 1157–1167.
7. Dabrowska, D. M. (1992). Nonparametric quantile regression with censored data. *Sankhyā* **54**, 252–259.
8. Delecroix, M., Härdle, W. & Hristache, M. (2003). Efficient estimation in conditional single-index regression. *J. Multivariate Anal.* **86**, 213–226.
9. Du, Y. & Akritas, M. G. (2002). I.i.d representations of the conditional Kaplan-Meier process for arbitrary distributions. *Math. Meth. Statist.* **11**, 152-182.
10. Einmahl, U. & Mason, D.M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33**, 1380–1403.
11. Gill, R.D. & Johansen, S. (1990). A Survey of Product-Integration with a View Toward Application in Survival Analysis. *Ann. Statist.* **18**, 1501–1555.
12. Hall, P. & Yao, Q. (2005). Approximating conditional distribution functions using dimension reduction. *Ann. Statist.* **33**, 1404–1421.
13. Horowitz, J. (2009). *Semiparametric and Nonparametric Methods in Econometrics.* Springer Series in Statistics. Springer: New-York.
14. Lavergne, P. & Patilea, V. (2013). Smooth Minimum Distance Estimation and Testing in Conditional Moment Restrictions Models: Uniform in Bandwidth Theory. *J. Econometrics* **177**, 47–59.

15. Lee, K.-Y., Li, B. & Chiaromonte, F. (2013). A general theory for nonlinear sufficient dimension reduction: Formulation and estimation. *Ann. Statist.* **41**, 221–249.

16. Li, G. & Datta S. (2001). Bootstrap approach to nonparametric regression for right censored data. *Ann. Inst. Statist. Math.* **53**, 708–729.

17. Li, W. & Patilea V. (2014). A new semiparametric estimator for single-index regressions. Working Paper, CREST-Ensai.

18. Lopez, O. (2011). Nonparametric estimation of the multivariate distribution function in a censored regression model with applications. *Comm. Statist. Theory Methods* **40**, 2639–2660.

19. Lu, X. (2010). Asymptotic distributions of two "syntheticdata" estimators for censored single-index models. *J. Multivariate Anal.* **101**, 999–1015.

20. Ma, Y. & Zhu, L. (2013). Efficient estimation in sufficient dimension reduction. *Ann. Statist.* **41**, 250–268.

21. Maistre, S. & Patilea, V. (2014). Nonparametric model checks of single-index assumptions. arXiv:1410.4934 [math.ST].

22. Miller, R. & Halpern, J. (1982). Regression with censored data. *Biometrika* **69**, 521–531.

23. Patilea, V., & Rolin, J.M. (2006). Product-limit estimators of the survival function for two modified forms of current-status data. *Bernoulli* **12**, 801-819.

24. Sherman, R.P. (1994). Maximal inequalities for degenerate $U-$processes with applications to optimization estimators. *Ann. Statist.* **22**, 439–459.

25. Strzalkowska-Kominiak, E. & Cao, R. (2013). Maximum likelihood estimation for conditional distribution single-index models under censoring. *J. Multivariate Anal.* **114**, 74–98.

26. Strzalkowska-Kominiak, E. & Cao, R. (2014). Beran-based approach for single-index models under censoring. *Comput. Statist.* **29**, 1243–1261.

27. Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal.* **45**, 89–103.

28. Tsodikov, A.D., J. G Ibrahim, J.G. & Yakovlev, A.Y. (2003). Estimating Cure Rates From Survival Data: An Alternative to Two-Component Mixture Models. *J. Amer. Statist. Assoc.* **98**, 1063–1078.

29. Van Keilegom, I. & Veraverbeke, N. (1996). Uniform strong convergence for the conditional Kaplan-Meier estimator and its quantiles. *Comm. Statist. Theory Methods* **25**, 2251-2265.

30. Van Keilegom, I. & Veraverbeke, N. (1997). Weak convergence of the bootstrapped conditional Kaplan-Meier process and its quantile process. *Comm. Statist. Theory Methods* **26**, 853–869.

31. Van Keilegom, I., Akritas, M.G. & Veraverbeke, N. (2001). Estimation of the conditional distribution in regression with censored data: a comparative study. *Comput. Statist. Data Anal.* **35**, 487–500.

32. Wei, L.J., Ying Z. & Lin, D.Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–851.

33. Xia, Y., Dixin Zhang, D. & Xu, J. (2010). Dimension Reduction and Semiparametric Estimation of Survival Models. *J Amer. Statist. Assoc.* **105**, 278–290.

34. Xu, J., & Peng, Y. (2014). Nonparametric cure rate estimation with covariates. *Canad.J. Statist.* **42**, 1–17.

35. Zheng, D., Yin, G. & Ibrahim, J.G. (2006). Semiparametric Transformation Models for Survival Data With a Cure Fraction. *J. Amer. Statist. Assoc.* **101**, 670–684.

IX Appendix

IX.1 Proof of equation (IV.3)

Here we justify the statement

$$I(\beta) \geq 0 \text{ and } I(\beta) = 0 \quad \text{if and only if} \quad \mathbb{E}[U(t,d;\beta) \mid X] = 0 \ \ a.s. \ \ \forall t \in \mathbb{R}, \ d \in \{0,1\}.$$

Let $\mathcal{F}[\omega](u) = \int_{\mathbb{R}^p} e^{-2\pi i x^\top u}\omega(x)dx$, $u \in \mathbb{R}^p$, denote the Fourier Transform of $\omega(\cdot)$. If $\mathcal{F}[\omega]$ is integrable, by the Inverse Fourier Transform formula and Fubini Theorem, we can write

$$\begin{aligned} I(\beta) &= \int_{\mathbb{R}\times\{0,1\}} \mathbb{E}\left[\omega(X_1 - X_2)U_1(t,d;\beta)U_2(t,d;\beta)\right] d\mu(t,d) \\ &= \int_{\mathbb{R}\times\{0,1\}} \mathbb{E}\left[U_1(t,d;\beta)U_2(t,d;\beta)\int_{\mathbb{R}^p} e^{2\pi i(X_1-X_2)^\top u}\mathcal{F}[\omega](u)du\right] d\mu(t,d) \\ &= \int_{\mathbb{R}\times\{0,1\}}\int_{\mathbb{R}^p} \left|\mathbb{E}\left[\mathbb{E}\left[U(t,d;\beta) \mid X\right] e^{2\pi i X^\top u}\right]\right|^2 \mathcal{F}[\omega](u)du d\mu(t,d). \end{aligned}$$

The statement follows from the fact that $\mathcal{F}[\omega]$ is positive and the uniqueness of the Fourier Transform.

IX.2 Assumptions

Let us introduce some notation. Let $\widetilde{X} \in \mathbb{R}^{p-1}$ be the $(p-1)-$dimension vector of the last components of X. Below, $(\widetilde{X})_r$ (resp. $(\widetilde{X}\widetilde{X}^\top)_{rq}$) denotes the rth components (resp. the $rq-$entry) of the vector $\widetilde{X}$ (resp. matrix $\widetilde{X}\widetilde{X}^\top$). If A is a matrix with real entries, $\|A\| = \sqrt{trace(A^\top A)}$. In the following, where ∂_z (resp. ∂^2_{zz}) denotes the first (resp. second) order derivative with respect to z. Some comments on the following assumptions are provided in the Supplementary Material.

Assumption IX.1. *1. The observations (Y_i, δ_i, X_i), $1 \leq i \leq n$, are independent copies of $(Y, \delta, X) \in \mathbb{R} \times \{0,1\} \times \mathbb{R}^p$. Moreover, there exists a positive number a such that $\mathbb{E}[\exp(a\|X\|)] < \infty$.*

2. The law of (Y, δ, X) is such that $\int_{\mathbb{R}}[H([t,\infty) \mid x)]^{-1}H_0(dt \mid x) = \infty$, $\forall x$.

3. *The parameter set is* $\mathcal{B} = \{1\} \times \mathcal{B}'$ *and* $\mathcal{B}' \subset \mathbb{R}^{p-1}$ *is a compact set with non-empty interior. The vector* $\beta_0 \in \mathcal{B}$ *satisfying the condition (III.1) is the unique element* $\mathcal{B}$ *having this property and the sub-vector built with its last* $p-1$ *components is in the interior of* $\mathcal{B}'$*. For any* $\beta \in \mathcal{B}$ *the random variable* $X^\top \beta$ *has a density* f_β*.*

4. *The value* β_0 *is a well-separated point of minimum for* $I(\beta)$ *defined in equation (IV.2) with* $\omega(x) = \exp(-\|x\|^2/2)$ *and* μ *equal to the distribution* $F_{Y,\delta}$ *of the observations* (Y, δ)*, that means, for any* $\varepsilon > 0$*,* $\inf_{\beta \in \mathcal{B}, \|\beta - \beta_0\| \geq \varepsilon} I(\beta) > I(\beta_0)$*.*

5. $$\sup_{z \in \mathbb{R}} \mathbb{E}\left[\|\widetilde{X}\|^4 \mid X^\top \beta_0 = z\right] f_{\beta_0}(z) < \infty \tag{IX.2}$$

6. *For each* $d \in \{0, 1\}$ *and* $t \in \mathbb{R}$*,* $0 \leq r, q \leq p-1$ *the functions*

$$z \mapsto f_{\beta_0}(z), \quad z \mapsto H_{d,\beta_0}((-\infty, t] \mid X^\top \beta_0 = z),$$

$$z \mapsto \mathbb{E}[(\widetilde{X})_r \mid X^\top \beta_0 = z] \quad \textit{and} \quad z \mapsto \mathbb{E}[(\widetilde{X}\widetilde{X}^\top)_{rq} \mid X^\top \beta_0 = z]$$

are four times continuously differentiable and the derivatives up to order four are bounded, respectively uniformly bounded with respect to t *in the case of* H_{d,β_0}*. The fourth order derivative are Lipschitz functions, with a Lipschitz constant independent of* t *in the case of* H_{d,β_0}*.*

7. *Let* A *be the set of values* $(t, d) \in \mathbb{R} \times \{0, 1\}$ *such that*

$$Var\left[\left(\widetilde{X} - \mathbb{E}\left[\widetilde{X} \mid X^\top \beta_0\right]\right) \partial_z \left(H_{d,\beta_0}((-\infty, t] \mid \cdot)\right)(X^\top \beta_0)\right] \tag{IX.3}$$

is positive definite. Then $F_{Y,\delta}(A) > 0$*.*

8. *Let* $z \mapsto \lambda_\beta(z; t, d)$ *denote any of the four functions at point 6 above and their derivatives up to the second order, considered for any* $\beta \in \mathcal{B}$*. Then, the family of functions* $\{\lambda_\beta(\cdot; t, d) : \beta \in \mathcal{B}, t \in \mathbb{R}, d = 0, 1\}$ *is a VC-class (or Euclidian) for an envelope with finite moment of order 8. Moreover, for any sequence* $b_n \to 0$*,*

$$\sup_{\|\beta - \beta_0\| \leq b_n} \sup_{t \in \mathbb{R}, d \in \{0,1\}} \sup_{z \in \mathbb{R}} |\lambda_\beta(z; t, d) - \lambda_{\beta_0}(z; t, d)| \to 0.$$

9. *The kernel $L(\cdot)$ is a symmetric and twice continuously differentiable univariate density with the second order derivative with bounded variation. Moreover, for $\kappa = 1, 2$, $\int_{\mathbb{R}} |L^{(\kappa)}(u)|du < \infty$, where $L^{(\kappa)}(\cdot)$ denotes the κth derivative of $L(\cdot)$.*

10. *$ng^4 \to 0$ and $ng^{3+a} \to \infty$ for some $a \in (0, 1)$.*

IX.3 Proof of Proposition 4.1

Let us introduce some notation. For $\beta \in \mathcal{B} \subset \{1\} \times \mathbb{R}^{p-1}$ let $\widetilde{\beta} \in \mathbb{R}^{p-1}$ denote the sub-vector of its last $(p-1)$ components. Let ∇_β the differential operator given by the $(p-1)$ last first order partial derivatives corresponding to the last $(p-1)$ components of β. This means, $\nabla_\beta \widehat{I}_n(\beta)$ represents the $(p-1)-$dimension vector of first order partial derivatives of $I_n(\beta)$ with respect to $\widetilde{\beta}$ and $\nabla^2_{\beta\beta} I_n(\beta)$ denotes the corresponding Hessian $(p-1) \times (p-1)-$matrix. For each i, let $\widetilde{X_i} \in \mathbb{R}^{p-1}$ be the $(p-1)-$dimension vector of the last components of X_i. We will also use the following simplified notation :

$$\sup_i = \sup_{1 \le i \le n}, \qquad \sup_{t,d} = \sup_{(t,d) \in \mathbb{R} \times \{0,1\}}.$$

In the following, for a sequence W_n, $n \ge 1$ of random vectors, $W_n = O_{\mathbb{P}}(1)$ (resp. $W_n = o_{\mathbb{P}}(1)$) means $\|W_n\| = O_{\mathbb{P}}(1)$ (resp. $\|W_n\| = o_{\mathbb{P}}(1)$).

Proof of Proposition 4.1. Since by assumption β_0 is a well-separated point of minimum for $I(\beta)$, it suffices to prove that

$$\sup_{\beta \in \mathcal{B}} \left| \widehat{I}_n(\beta) - I(\beta) \right| = o_{\mathbb{P}}(1).$$

This uniform convergence follows by Hoeffding decomposition of $U-$statistic of order 4 obtained from $\widehat{I}_n(\beta)$ after removing negligible diagonal terms, and by results on the rate of uniform convergence of degenerate $U-$statistics, like for instance those of Sherman (1994). See also the Maximal Inequality in Supplementary Material. Li & Patilea (2014) provide complete arguments in the case where $\mathbb{P}(\delta = 1) = 1$, our case herein is very similar and hence we omit the details.

First, we prove that $\widehat{\beta} - \beta_0 = O_{\mathbb{P}}(n^{-1/2})$. For this purpose, it suffices to prove

$$\nabla_\beta \widehat{I}_n(\beta_0) = O_{\mathbb{P}}(n^{-1/2}) \tag{IX.4}$$

and there exists a positive definite matrix $J(\beta_0)$ such that for any sequence of $\overline{\beta}$ between $\widehat{\beta}$ and β_0,

$$\mathbb{P}\left(1/c \leq \|\nabla^2_{\beta\beta} \widehat{I}_n(\overline{\beta}) - J(\beta_0)\| \leq c\right) \to 1, \tag{IX.5}$$

for some $c > 0$. See for instance Theorem 1-(ii) of Sherman (1994).

Let us write

$$\widehat{U}_i(t, d; \beta) = \frac{1}{n-1} \sum_{k=1}^{n} \left\{ \mathbf{1}\{Y_i \leq t; \delta_i = d\} - \mathbf{1}\{Y_k \leq t; \delta_k = d\} \right\} \frac{1}{g} L_{ik}(\beta, g),$$

where $L_{ik}(\beta, g) = L((X_i - X_k)^\top \beta / g)$, and recall the notation

$$U_i(t, d; \beta) = \left\{ \mathbf{1}\{Y_i \leq t, \delta_i = d\} - H_{d,\beta}((-\infty, t] \mid X_i^\top \beta) \right\} f_\beta(X_i^\top \beta).$$

If

$$\widehat{I}_n(t, d; \beta) = \frac{1}{n^2} \sum_{1 \leq i,j \leq n} \widehat{U}_i(t, d; \beta) \widehat{U}_j(t, d; \beta) \omega_{ij}, \qquad (t, d) \in \mathbb{R} \times \{0, 1\},$$

with $\omega_{ij} = \omega(X_i - X_j)$, then

$$\begin{aligned} \frac{1}{2} \nabla_\beta \widehat{I}_n(t, d; \beta_0) &= \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \nabla_\beta \widehat{U}_i(t, d; \beta_0) \widehat{U}_j(t, d; \beta_0) \omega_{ij} \\ &\quad + \frac{\omega(0)}{n^2} \sum_{1 \leq i \leq n} \nabla_\beta \widehat{U}_i(t, d; \beta_0) \widehat{U}_i(t, d; \beta_0) \\ &= D_{n1}(t, d; \beta_0) + D_{n2}(t, d; \beta_0). \end{aligned}$$

Since $\widehat{I}_n(\beta_0)$ is the integral of $\widehat{I}_n(\cdot, \cdot; \beta_0)$ with respect to $\widehat{F}_{n,Y,\delta}$, the empirical distribution of the observations (Y, δ), the rate of the former could be derived from the uniform rate of the latter. In the following, when there is no danger of confusion, we simplify the writings and omit the arguments t, d, β_0. We will focus on the uniform rate of D_{n1} since the arguments for D_{n2}, which is uniformly negligible, are similar

and much simpler. For this purpose we decompose $\widehat{U}_i$ and $\nabla_\beta \widehat{U}_i$ as follows:

$$\widehat{U}_i = \left\{\widehat{U}_i - \mathbb{E}\left[\widehat{U}_i \mid Y_i, \delta_i, X_i\right]\right\} + \left\{\mathbb{E}\left[\widehat{U}_i \mid Y_i, \delta_i, X_i\right] - U_i\right\} + U_i = V_{U,i} + B_{U,i} + U_i$$

and

$$\begin{aligned}\nabla_\beta \widehat{U}_i &= \left\{\nabla_\beta \widehat{U}_i - \mathbb{E}\left[\nabla_\beta \widehat{U}_i \,|\, X_i\right]\right\} + \left\{\mathbb{E}\left[\nabla_\beta \widehat{U}_i \,|\, X_i\right] - \mathbb{E}\left[\nabla_\beta U_i \,|\, X_i\right]\right\} + \mathbb{E}\left[\nabla_\beta U_i \,|\, X_i\right] \\ &= V_{\nabla,i} + B_{\nabla,i} + \mathbb{E}\left[\nabla_\beta U_i \mid X_i\right].\end{aligned}$$

By Lemma X.1,

$$\sup_i \sup_{t,d} \left\{|B_{U,i}| + \|B_{\nabla,i}\|\right\} = O_{\mathbb{P}}(g^2),$$

and for any $\alpha \in (0,1)$,

$$\sup_i \sup_{t,d} \left\{|V_{U,i}| + g\,\|V_{\nabla,i}\|\right\} = O_{\mathbb{P}}(n^{-1/2} g^{\alpha/2-1}).$$

Then we can decompose

$$\begin{aligned}\frac{n-1}{n} D_{n1}(t,d;\beta_0) &= \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \mathbb{E}\left[\nabla_\beta U_i \mid X_i\right] \left\{U_j + V_{U,j}\right\} \omega_{ij} \\ &\quad + \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} V_{\nabla,i}\{V_{U,j} + U_j\}\, \omega_{ij} \\ &\quad + O_{\mathbb{P}}(g^2) + O_{\mathbb{P}}\left(n^{-1/2} g^{\alpha/2}\right).\end{aligned}$$

By Lemma X.2,

$$\sup_{t,d} \left\| \frac{1}{n(n-1)} \sum_{i\ne j} \mathbb{E}\left[\nabla_\beta U_i \mid X_i\right] \left\{U_j + V_{U,j}\right\} \omega_{ij} \right\| = O_{\mathbb{P}}\left(n^{-1/2}\right).$$

Moreover,

$$\sup_{t,d} \frac{1}{n(n-1)} \left\| \sum_{i\ne j} V_{\nabla,i}\{U_j + V_{U,j}\}\, \omega_{ij} \right\| = o_{\mathbb{P}}\left(n^{-1/2}\right).$$

Thus, integrating $D_{n1}(t,d;\beta_0)$ with respect to the empirical distribution function of the (Y_i, δ_i)'s yields the rate $O_{\mathbb{P}}(n^{-1/2})$. From similar arguments applied for $D_{n2}(t,d;\beta_0)$ one deuces the rate (IX.4).

For the second order derivative we can decompose

$$\begin{aligned}\frac{1}{2}\nabla^2_{\beta\beta}\widehat{I}_n(t,d;\beta) &= \frac{1}{n^2}\sum_{1\le i,j\le n}\nabla_\beta\widehat{U}_i(t,d;\beta)\nabla_\beta\widehat{U}_j(t,d;\beta)^\top\omega_{ij}\\ &+ \frac{1}{n^2}\sum_{1\le i,j\le n}\nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta)\widehat{U}_j(t,d;\beta)\omega_{ij}\\ &= E_{n1}(t,d;\beta)+E_{n2}(t,d;\beta).\end{aligned}$$

It is shown in Lemma X.4 in the Supplementary Material that, for any sequence $b_n \to 0$,

$$\sup_{t\in\mathbb{R},\, d\in\{0,1\}}\sup_{\|\beta-\beta_0\|\le b_n}|E_{n1}(t,d;\beta)-E_{n1}(t,d;\beta_0)| = o_{\mathbb{P}}(1)$$

and

$$\sup_{t\in\mathbb{R},\, d\in\{0,1\}}\sup_{\|\beta-\beta_0\|\le b_n}|E_{n2}(t,d;\beta)-E_{n2}(t,d;\beta_0)| = o_{\mathbb{P}}(1).$$

Here, the sequence (b_n) tending to zero is such that $\mathbb{P}[\|\widehat{\beta}-\beta_0\|\le b_n]\to 1$. On the other hand, by Lemmas X.1 and X.3 in the Supplementary Material, for any $\alpha\in(0,1)$,

$$\begin{aligned}E_{n2}(t,d;\beta_0) &= \frac{1}{n^2}\sum_{1\le i,j\le n}\left\{V_{\nabla^2,i}+B_{\nabla^2,i}+\mathbb{E}\left[\nabla^2_{\beta\beta}U_i\mid X_i\right]\right\}\left\{V_{U,j}+B_{U,j}+U_j\right\}\omega_{ij}\\ =&\frac{1}{n^2}\sum_{1\le i,j\le n}V_{\nabla^2,i}U_j\omega_{ij}+\frac{1}{n^2}\sum_{1\le i,j\le n}\left\{O_{\mathbb{P}}(n^{-1/2}g^{\alpha/2-3})\right\}\left\{O_{\mathbb{P}}(n^{-1/2}g^{\alpha/2-1})+O_{\mathbb{P}}(g^2)\right\}\omega_{ij}\\ &+\frac{1}{n^2}\sum_{1\le i,j\le n}\left\{O_{\mathbb{P}}(g^2)\right\}\left\{O_{\mathbb{P}}(n^{-1/2}g^{\alpha/2-1})+O_{\mathbb{P}}(g^2)+U_j\right\}\omega_{ij}\\ +&\frac{1}{n^2}\sum_{1\le i,j\le n}\{O_{\mathbb{P}}(1)\}\left\{O_{\mathbb{P}}(n^{-1/2}g^{\alpha/2-1})+O_{\mathbb{P}}(g^2)\right\}\omega_{ij}+\frac{1}{n^2}\sum_{1\le i,j\le n}\mathbb{E}\left[\nabla^2_{\beta\beta}U_i\mid X_i\right]U_j\omega_{ij}.\end{aligned}$$

Given the uniform rates of the first and the last terms in the previous decomposition, see Lemma X.3, the fact that α could be arbitrarily close to 1, and the fact that

$ng^{3+a} \to \infty$ for some $a \in (0,1)$, deduce that

$$\sup_{t\in\mathbb{R},\, d\in\{0,1\}} |E_{n2}(t,d;\beta_0)| = o_{\mathbb{P}}(1)$$
$$+ \sup_{t\in\mathbb{R},\, d\in\{0,1\}} \left\{ \left| \frac{1}{n^2} \sum_{1\le i,j\le n} V_{\nabla^2,i} U_j \omega_{ij} \right| + \left| \frac{1}{n^2} \sum_{1\le i,j\le n} \mathbb{E}\left[\nabla^2_{\beta\beta} U_i \mid X_i \right] U_j \omega_{ij} \right| \right\}$$
$$= o_{\mathbb{P}}(1).$$

It remains to investigate the uniform convergence of the term $E_{n1}(t,d;\beta_0)$. By Lemmas X.1 and X.2,

$$\sup_{t\in\mathbb{R},\, d\in\{0,1\}} \left| E_{n1}(t,d;\beta_0) - \mathbb{E}\Big\{ \mathbb{E}\left[\nabla_\beta U_1(t,d;\beta_0)|X_1\right] \mathbb{E}\Big[\nabla_\beta U_2(t,d;\beta_0)^\top | X_2\Big] \omega_{12} \Big\} \right| = o_{\mathbb{P}}(1).$$

Thus

$$\int \Big[E_{n1}(t,d;\beta_0) - \mathbb{E}\Big[\mathbb{E}[\nabla_\beta U_1(t,d;\beta_0) \,|\, X_1] \mathbb{E}[\nabla_\beta U_2(t,d;\beta_0)^\top \,|\, X_2] \omega_{12} \Big] \Big] d\widehat{F}_{n,Y,\delta}(t,d) \to 0,$$

where $\widehat{F}_{n,Y,\delta}$ is the empirical distribution of the observations (Y_i, δ_i) with true distribution function $F_{Y,\delta}$. As a consequence, by the law of large numbers,

$$\int E_{n1}(t,d;\beta_0) d\widehat{F}_{n,Y,\delta}(t,d) - J(\beta_0) = o_{\mathbb{P}}(1),$$

where

$$J(\beta_0) = \int \mathbb{E}\left\{ \mathbb{E}[\nabla_\beta U_1(t,d;\beta_0) \mid X_1] \mathbb{E}[\nabla_\beta U_2(t,d;\beta_0)^\top \mid X_2] \omega_{12} \right\} dF_{Y,\delta}(t,d). \quad \text{(IX.6)}$$

En explicit expression of $\mathbb{E}[\nabla_\beta U_i(t,d;\beta_0) \mid X_i]$ is provided in equation (X.3) in the Supplementary Material. Our Assumption IX.1-7 guarantees that $J(\beta_0)$ is positive definite. Gathering rates, deduce that the property (IX.5) holds true and thus $\widehat{\beta}$ is $\sqrt{n}-$consistent.

For the asymptotic normality, by the definition of $\widehat{\beta}$ and a first order Taylor expansion,

$$0 = \nabla_\beta \widehat{I}_n(\widehat{\beta}) = \nabla_\beta \widehat{I}_n(\beta_0) + \nabla^2_{\beta\beta} \widehat{I}_n(\overline{\beta})(\widetilde{\widehat{\beta}} - \widetilde{\beta}_0),$$

where $\overline{\beta}$ is a vector between $\widehat{\beta}$ and β_0. Next, it suffices to follow the proof of the $\sqrt{n}-$asymptotic normality and deduce that

$$\nabla_\beta \widehat{I}_n(\beta_0) = 2\int \mathcal{V}_n(t,d;\beta_0) dF_{n,Y,\delta}(t,d) + o_{\mathbb{P}}(n^{-1/2}) \quad \text{and} \quad \nabla^2_{\beta\beta}\widehat{I}_n(\overline{\beta}) = J(\beta_0) + o_{\mathbb{P}}(1),$$

for any sequence $\overline{\beta}$ between $\widehat{\beta}$ and β_0, and

$$\mathcal{V}_n(t,d;\beta_0) = \frac{1}{n(n-1)} \sum_{i\neq j} \mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right] U_j(t,d;\beta_0)\, \omega_{ij}.$$

One can decompose

$$\begin{aligned}\mathcal{V}_n(t,d;\beta_0) &= \frac{1}{n}\sum_{1\leq j\leq n} \mathbb{E}\left\{\mathbb{E}\left[\nabla_\beta U(t,d;\beta_0) \mid X\right]\omega(X-X_j) \mid X_j\right\} U_j(t,d;\beta_0) \\ &+ \frac{1}{n(n-1)}\sum_{i\neq j}\left[\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0)\,|\,X_i\right]\omega_{ij} - \mathbb{E}\{\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0)\,|\,X_i\right]\omega_{ij}\,|\,X_j\}\right] U_j(t,d;\beta_0) \\ &= \mathcal{V}_{1n}(t,d;\beta_0) + \mathcal{V}_{2n}(t,d;\beta_0).\end{aligned}$$

The maximal inequality (X.2) implies that the degenerate second order $U-$process $\mathcal{V}_{2n}$ is of uniform rate $O_{\mathbb{P}}(n^{-1})$, and hence negligible. On the other hand,

$$\int \mathcal{V}_{1n}(t,d;\beta_0) dF_{n,Y,\delta}(t,d) = \int \mathcal{V}_{1n}(t,d;\beta_0) dF_{Y,\delta}(t,d) + o_{\mathbb{P}}(n^{-1/2}),$$

as shown in Lemma X.7 in the Supplementary Material. Finally, the multivariate Central Limit Theorem implies that $\int \mathcal{V}_{1n}(t,d;\beta_0) dF_{Y,\delta}(t,d)$ is asymptotically normal and thus

$$\sqrt{n}\nabla_\beta \widehat{I}_n(\beta_0) \rightsquigarrow N(0,\Sigma(\beta_0)),$$

where

$$\Sigma(\beta_0) = 4\mathbb{E}\left[\psi(Y,\delta,X;\beta_0)\psi(Y,\delta,X;\beta_0)^\top\right], \tag{IX.7}$$

is a positive definite $(p-1)\times(p-1)-$matrix and

$$\psi(Y_j,\delta_j,X_j;\beta_0) = \int \mathbb{E}\left\{\mathbb{E}\left[\nabla_\beta U(t,d;\beta_0) \mid X\right]\omega(X-X_j) \mid X_j\right\} U_j(t,d;\beta_0) dF_{Y,\delta}(t,d).$$

Finally deduce

$$\sqrt{n}(\widehat{\beta} - \beta_0) \rightsquigarrow N(0, V(\beta_0)) \qquad \text{where} \qquad V(\beta_0) = J(\beta_0)^{-1}\Sigma(\beta_0)J(\beta_0)^{-1},$$

and $J(\beta_0)$ and $\Sigma(\beta_0)$ are defined as in equations (IX.6) and (IX.7). Now the proof of Proposition 4.1 is complete. □

IX.4 Proof of Proposition 5.1

The idea of the proof is to show that

$$\sup_{x\in\overline{\mathcal{X}}} \sup_{t\leq\tau} \left|\widehat{\Lambda}_{T,\widehat{\beta}}((-\infty, t] \mid x^\top\widehat{\beta}) - \widehat{\Lambda}_{T,\beta_0}((-\infty, t] \mid x^\top\beta_0)\right| = o_{\mathbb{P}}(n^{-1/2}h^{-1/2}), \quad \text{(IX.8)}$$

and then to apply the i.i.d. representations of Du & Akritas (2002) and Lopez (2011) for $\widehat{\Lambda}_{T,\beta_0}((-\infty, t] \mid x^\top\beta_0)$. The representation for $\widehat{F}_{T,\beta_0}((-\infty, t] \mid x^\top\beta_0)$ will follow easily. Since for any given β

$$\widehat{\Lambda}_{T,\beta}((-\infty, t] \mid x^\top\beta) = \frac{1}{nh}\sum_{i=1}^{n} \frac{\mathbf{1}\{Y_i \leq t\}\delta_i K((X_i - x)^\top\beta/h)}{\frac{1}{nh}\sum_{k=1}^{n}\mathbf{1}\{Y_k \geq Y_i\}K((X_k - x)^\top\beta/h)},$$

it suffices, on one hand, to show that

$$\sup_{\|\beta-\beta_0\|\leq b_n} \sup_{t\leq\tau} \sup_{x\in\overline{\mathcal{X}}} |\Delta_n(t, x; \beta)| = o_{\mathbb{P}}(n^{-1/2}h^{-1/2}),$$

where

$$\Delta_n(t, x; \beta) = \frac{1}{nh}\sum_{i=1}^{n} \xi(Y_i, \delta_i; t)\left[K((X_i - x)^\top\beta/h) - K((X_i - x)^\top\beta_0/h)\right],$$

with

$$\xi(Y_i, \delta_i; t) = \mathbf{1}\{Y_i \leq t\}\delta_i \quad \text{or} \quad \xi(Y_i, \delta_i; t) = \mathbf{1}\{Y_i \geq t\}, \qquad t \leq \tau,$$

and $b_n = O_{\mathbb{P}}(n^{-1/2})$. On the other hand, to guarantee that, for some constant $c > 0$,

$$D_n = \inf_{x\in\overline{\mathcal{X}}} \frac{1}{nh}\sum_{k=1}^{n} \mathbf{1}\{Y_k \geq \tau\}K((X_k - x)^\top\beta_0/h) \geq c$$

with probability tending to 1.

To uniformly bound Δ_n let us write

$$\begin{aligned}|\Delta_n(t,x;\beta)| &\le |\Delta_{1n}(t,x;\beta)-\Delta_{1n}(t,x;\beta_0)| \\ &\quad + \left|\mathbb{E}\left[\xi(Y,\delta;t)h^{-1}\left\{K((X-x)^\top\beta/h)-K((X-x)^\top\beta_0/h)\right\}\right]\right| \\ &= |\Delta_{1n}(t,x;\beta)-\Delta_{1n}(t,x;\beta_0)| + |\Delta_{2n}(t,x;\beta,\beta_0)|\,,\end{aligned}$$

where

$$\Delta_{1n}(t,x;\beta)=\frac{1}{nh}\sum_{i=1}^n\left\{\xi(Y_i,\delta_i;t)K((X_i-x)^\top\beta/h)-\mathbb{E}\left[\xi(Y,\delta;t)K((X-x)^\top\beta/h)\right]\right\}.$$

For any fixed β, the process $\Delta_{1n}(t,x;\beta)$ indexed by $t\le\tau$ and $x\in\overline{\mathcal{X}}$ is of uniform rate $O_{\mathbb{P}}(n^{-1/2}h^{-1/2}\log^{1/2}n)$. Letting β to approach β_0 at the rate $O_{\mathbb{P}}(n^{-1/2})$, we could use the modulus of continuity of the process $\Delta_{1n}(t,x;\beta)$ indexed by t, x and β and thus derive the required uniform rate $o_{\mathbb{P}}(n^{-1/2}h^{-1/2})$ for $\Delta_{1n}(t,x;\beta)-\Delta_{1n}(t,x;\beta_0)$. By Taylor expansion, suitable changes of variable and the assumptions guaranteeing that $|f_\beta(\cdot)-f_{\beta_0}(\cdot)|/\|\beta-\beta_0\|$ is bounded and $f_{\beta_0}(\cdot)$ is Lipschitz continuous, $\Delta_{2n}(t,x;\beta,\beta_0)$ could be also shown to be of uniform rate $o_{\mathbb{P}}(n^{-1/2}h^{-1/2})$. The details are provided in Lemma X.6 in the Supplementary Material.

To bound D_n, let us decompose

$$\begin{aligned}D_n &= \inf_{x\in\overline{\mathcal{X}}}\mathbb{E}\{H_{\beta_0}([\tau,\infty)\mid X^\top\beta_0)h^{-1}K((X-x)^\top\beta_0/h)\} \\ &-\sup_{x\in\overline{\mathcal{X}}}\frac{1}{nh}\sum_{k=1}^n\Big[\mathbf{1}\{Y_k\ge\tau\}K((X_k-x)^\top\beta_0/h)-\mathbb{E}\{H_{\beta_0}([\tau,\infty)\,|\,X^\top\beta_0)K((X-x)^\top\beta_0/h)\}\Big] \\ &= D_{1n}-D_{2n}.\end{aligned}$$

By standard results in the uniform convergence of kernel estimators, see for instance Theorem 4 of Einmahl & Mason (2005), $D_{2n}=O_{\mathbb{P}}(n^{-1/2}h^{-1/2}\log^{1/2}n)=o_{\mathbb{P}}(1)$. (A slightly slower uniform rate, but still negligible, could be obtained using the Maximal Inequality of Sherman (1994) recalled in the Supplementary Material.) It remains to

show that D_{1n} stays away from zero. By a change of variables and Taylor expansion,

$$\mathbb{E}\{H_{\beta_0}([\tau,\infty) \mid X^\top\beta_0)h^{-1}K((X-x)^\top\beta_0/h)\} = H_{\beta_0}([\tau,\infty) \mid x^\top\beta_0)f_{\beta_0}(x^\top\beta_0)+O_{\mathbb{P}}(h^2),$$

uniformly with respect to $x \in \overline{\mathcal{X}}$. By construction and assumptions,

$$\inf_{x\in\overline{\mathcal{X}}} H_{\beta_0}([\tau,\infty) \mid x^\top\beta_0)f_{\beta_0}(x^\top\beta_0) > 0,$$

so that D_n is uniformly bounded away from zero. Gathering facts, one deduces the uniform rate in equation (IX.8). Now, by the Duhamel identity, see Gill & Johansen (1990), page 1635, one deduces

$$\sup_{x\in\overline{\mathcal{X}}}\sup_{t\leq\tau}\left|\widehat{F}_{T,\widehat{\beta}}((-\infty,t] \mid x^\top\widehat{\beta}) - \widehat{F}_{T,\beta_0}((-\infty,t] \mid x^\top\beta_0)\right| = o_{\mathbb{P}}(n^{-1/2}h^{-1/2}).$$

See also Xu & Peng (2014), page 14. Then the representation for $\widehat{F}_{T,\widehat{\beta}}$ follows from the i.i.d. representation of the conditional survival function derived by Du & Akritas (2002) and Lopez (2011). Now the proof is complete.

IX.5 Proof of Corollary 5.2

Let $\tau = \sup_{x\in\mathcal{X}} \tau_{H_1}(x;\beta_0)$. By definition $Y^1_{(n)} \leq \tau$ and

$$\widehat{F}_{T,\widehat{\beta}}((Y^1_{(n)},\infty]) \mid x^\top\widehat{\beta}) = \widehat{F}_{T,\widehat{\beta}}((\tau,\infty]) \mid x^\top\widehat{\beta}), \qquad \forall x \in \mathcal{X}.$$

Next, by assumption we have $\tau < \inf_{x\in\overline{\mathcal{X}}} \tau_H(x;\beta_0)$. Since obviously

$$\pi(x) = F_{T,\beta_0}((\tau_H(x),\infty] \mid x^\top\beta_0), \qquad \forall x \in \mathcal{X},$$

the $\sqrt{nh}-$asymptotic normality follows directly from the i.i.d. representation (V.8) and a central limit theorem applied for the i.i.d. sum $n^{-1}\sum_{i=1}^n \eta_{F_T,i}(\tau, x^\top\beta_0)$ considered with some $x \in \overline{\mathcal{X}}$.

X Supplementary material: comments, technical lemmas and proofs

In the following $C, C_1, C_2, C', \cdots$ represent constants, independent of the sample size. Their value may change from line to line. For nonnegative quantities a_n and r_n possibly depending on n, we will use the notation $a_n \lesssim r_n$ to indicate that there exists some constant C such that for each n, $a_n \leq C r_n$.

X.1 Comments on the Assumptions IX.1

The exponential moment imposed by condition 1 implies that the largest value of the sample $\|X_1\|, \cdots, \|X_n\|$ has the rate $O_{\mathbb{P}}(\log n)$. However, X is not required to be bounded or to have a density. The condition 2 guarantees that $F_{C,\beta}(\mathbb{R} \mid z) = 1$, $\forall z \in \mathbb{R}$, $\forall \beta \in \mathcal{B}$. The condition 4 is a classical condition used to prove consistency. The regularity imposed by condition 6, together with the Taylor expansion, serve to show that biases are negligible. In order to have a nondegenerate limit for $\widehat{\beta}$ we need the matrix

$$\int \mathbb{E}\left[\mathbb{E}[\nabla_\beta U_1(t,d;\beta_0) \mid X_1]\mathbb{E}[\nabla_\beta U_2(t,d;\beta_0)^\top \mid X_2]\omega_{12}\right] dF_{Y,\delta}(t,d) \tag{X.1}$$

to be definite positive. This means, if $v \in \mathbb{R}^{d-1}$ satisfies

$$\int \mathbb{E}\left[v^\top \mathbb{E}[\nabla_\beta U_1(t,d;\beta_0) \mid X_1]\mathbb{E}[\nabla_\beta U_2(t,d;\beta_0)^\top \mid X_2] v\, \omega_{12}\right] dF_{Y,\delta}(t,d) = 0,$$

then necessarily $v = 0$. By construction, the integrand in the last display is nonnegative. For any t, d such that the integrand is positive, by the Inverse Fourier Transform and the fact that $\mathcal{F}[\omega] > 0$, deduce that $\mathbb{E}[v^\top \nabla_\beta U(t,d;\beta_0) \mid X] = 0$ almost surely. This means that the variance of $\mathbb{E}[v^\top \nabla_\beta U(t,d;\beta_0) \mid X]$ is zero. Then, the condition of positive definiteness of the variance defined in (IX.3) for a set of values (t,d) of positive probability implies that necessarily $v = 0$. Condition 8 is used to control the difference of the second order derivative $\nabla^2_{\beta\beta}\widehat{I}_n(\beta) - \nabla^2_{\beta\beta}\widehat{I}_n(\beta_0)$ when $\beta - \beta_0$ tends to zero. The properties of the kernel $L(\cdot)$ allow for the Taylor

expansion used to study the bias of various quantities and to guarantee the VC-class property for the families indexing various $U-$processes appearing in the proof. (The definition of VC-class is recalled in the Supplementary Material.) The bandwidth g should be small enough in order to make bias terms negligible. This explains the condition $ng^4 \to 0$. The other condition of g is a convenient restriction to control variance terms.

X.2 Technical lemmas and proofs

Let $\mathcal{S}$ be an arbitrary space where the i.i.d. observations take value. Let m be a positive integer and $\mathcal{F}$ a class of real-valued functions on the product space $\mathcal{S}^m = \mathcal{S} \otimes \cdots \otimes \mathcal{S}$. Let F be an envelope for $\mathcal{F}$, that is $\sup_{\mathcal{F}} |f(\cdot)| \leq F(\cdot)$. For ν a probability measure on $\mathcal{S}^m$ and $\varepsilon > 0$, let $N(\varepsilon \|F\|_{L^2(\nu)}, \mathcal{F}, L^2(\nu))$, denote the covering number, that is the minimal number of balls of radius $\varepsilon \|F\|_{L^2(\nu)}$ in $L^2(\nu)$ needed to cover $\mathcal{F}$. See van der Vaart and Wellner (1996) or Sherman (1994) for the definitions. To derive our main results, we will need to derive the rates of uniform convergence for degenerate $U-$processes of order m indexed by VC (or Euclidian) classes of functions $\mathcal{F}$. A VC-class with constants (A, V) is a class of functions with covering number bounded by the polynomial $A\varepsilon^{-V}$, $\forall 0 < \varepsilon \leq 1$. A $U-$processes of order m indexed by $\mathcal{F}$, denoted by $U_n^m f$, is degenerate if for each $f \in \mathcal{F}$, $\int_{\mathcal{S}} f(s_1, \ldots, s_{j-1}, \cdot, s_{j+1}, \ldots, s_m) dP = 0$, $j = 1, \ldots, m$, where P is the probability distribution of the observations. The case $m = 1$ corresponds to empirical processes. Let us recall a simplified version of a general result of Sherman (1994) that could apply in all cases we need to investigate below. It is certainly not the sharpest result of this kind but, in particular, it allows to work with squared integrable envelopes so that is no longer necessary to truncate the functions from $\mathcal{F}$ and treat the tails separately. Hence, for the purposes of our theoretical study, we consider the uniform bound below a readable compromise that allows for only little loss of generality in the conditions on the bandwidth g. For the classes of functions we consider, the VC-class property is a direct consequence of standard results available, for instance, in Nolan and Pollard (1987), Pakes and Pollard (1989), Sherman (1994), van der Vaart and Wellner (1996). Hence, in following we will omit the details for

justifying the VC-class property.

Maximal Inequality [Sherman (1994), Main Corollary] *Let $\mathcal{F}$ be a class of degenerate functions on $\mathcal{F}$ on $\mathcal{S}^m$, $m \geq 1$. Suppose $\mathcal{F}$ is a VC-class of parameters (A, V) for a squared integrable envelope F. Then, for any $\alpha \in (0,1)$,*

$$\mathbb{E} \sup_{\mathcal{F}} \left| n^{m/2} U_n^p f \right| \leq \Lambda \left[\mathbb{E} \sup_{\mathcal{F}} \left(U_{2n}^m f^2 \right)^{\alpha} \right]^{1/2} \tag{X.2}$$

with $\Lambda > 0$ a constant depending on A, V, m and α, but independent of n.

Lemma X.1. *Under the condition of Proposition 4.1:*

1. $$\sup_i \sup_{t,d} \{ |B_{U,i}| + \|B_{\nabla,i}\| \} = O_{\mathbb{P}}(g^2).$$

2. *For any $\alpha \in (0,1)$,*

$$\sup_i \sup_{t,d} \{ |V_{U,i}| + g \, \|V_{\nabla,i}\| \} = O_{\mathbb{P}}(n^{-1/2} g^{\alpha/2-1}).$$

Proof of Lemma X.1. 1. For the uniform rate on the bias $B_{U,i}$ it suffices to show that

$$\mathbb{E}\left[g^{-1} L((X - X_i)^\top \beta_0 / g) \mid X_i \right] - f_{\beta_0}(X_i^\top \beta_0) = O_{\mathbb{P}}(g^2)$$

and

$$\mathbb{E}\left[H_{d,\beta_0}((-\infty, t] \,|\, X^\top \beta_0) g^{-1} L((X - X_i)^\top \beta_0 / g) \,|\, X_i \right] - H_{d,\beta_0}((-\infty, t] \,|\, X_i^\top \beta_0) f_\beta(X_i^\top \beta_0) = O_{\mathbb{P}}(g^2),$$

uniformly with respect to i, t and d. These uniform rates follow by standard change of variables and Taylor expansion of the functions $z \to f_{\beta_0}(z)$ and $z \to H_{d,\beta}((-\infty, t] \mid z) f_{\beta_0}(z)$, which, by our assumptions, have the required regularity.

For the uniform rate on the bias $B_{\nabla^2,i}$ let us write

$$\begin{aligned}\mathbb{E}\left[\nabla_\beta \widehat{U}_i(t,d;\beta_0) \mid X_i\right] &= H_{d,\beta_0}((-\infty,t] \mid X_i^\top\beta_0) \\ &\quad \times \mathbb{E}\left[\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X}_k \,|\, X_k^\top\beta_0]\right) g^{-2}L'((X_i - X_k)^\top\beta_0/g) \,|\, X_i\right] \\ &- \mathbb{E}\left[H_{d,\beta_0}((-\infty,t] \mid X_k^\top\beta_0)\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X}_k \mid X_k^\top\beta_0]\right)\right. \\ &\quad \left. \times g^{-2}L'((X_i - X_k)^\top\beta_0/g) \mid X_i\right].\end{aligned}$$

We used the fact that for each t and d, by the single-index assumption on the law of (Y,δ), we have $H_d((-\infty,t] \mid X_i) = H_{d,\beta_0}((-\infty,t] \mid X_i^\top\beta_0)$ and

$$\mathbb{E}\left[\widetilde{X}_k \mathbf{1}\{Y_k \le t, \delta_k = d\} \mid X_k^\top\beta_0\right] = H_{d,\beta_0}((-\infty,t] \mid X_k^\top\beta_0)\mathbb{E}[\widetilde{X}_k \mid X_k^\top\beta_0].$$

(Recall that $\widetilde{X}_i \in \mathbb{R}^{p-1}$ denotes the $(p-1)-$dimension vector of the last components of X_i.) Next, by integration by parts and Taylor expansion

$$\begin{aligned}&\mathbb{E}\left[\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X}_k \mid X_k^\top\beta_0]\right) g^{-2}L'((X_i - X_k)^\top\beta_0/g) \mid X_i\right] \\ &\quad = \int \partial_z\left(\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X} \mid X^\top\beta_0 = \cdot]\right) f_{\beta_0}(\cdot)\right)(z) g^{-1}L((X_i^\top\beta_0 - z)/g)dz \\ &\quad = \partial_z\left(\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X} \mid X^\top\beta_0 = \cdot]\right) f_{\beta_0}(\cdot)\right)(X_i^\top\beta_0) + O_{\mathbb{P}}(g^2),\end{aligned}$$

where for any univariate map $z \mapsto \zeta(z)$, $\partial_z\zeta$ denotes its derivative. Similarly,

$$\begin{aligned}&\mathbb{E}\left[H_{d,\beta_0}((-\infty,t] \mid \cdot)\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X}_k \mid X_k^\top\beta_0]\right) g^{-2}L'((X_i - X_k)^\top\beta_0/g) \mid X_i\right] \\ &= \int \partial_z\left(H_{d,\beta_0}((-\infty,t] \mid \cdot)\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X} \mid X^\top\beta_0 = \cdot]\right) f_{\beta_0}(\cdot)\right)(z) g^{-1}L((X_i^\top\beta_0 - z)/g)dz \\ &\quad = \partial_z\left(H_{d,\beta_0}((-\infty,t] \mid \cdot)\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X} \mid X^\top\beta_0 = \cdot]\right) f_{\beta_0}(\cdot)\right)(X_i^\top\beta_0) + O_{\mathbb{P}}(g^2).\end{aligned}$$

By our assumptions, the last two $O_{\mathbb{P}}(g^2)$ rates above are uniform with respect to i, t and d. Deduce that

$$\sup_i \sup_{t,d} \left\|\mathbb{E}\left[\nabla_\beta \widehat{U}_i(t,d;\beta_0) \mid X_i\right] - \mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right]\right\| = O_{\mathbb{P}}(g^2),$$

where

$$\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right] = -\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X}_i \mid X_i^\top\beta_0]\right) \times f_{\beta_0}(X_i^\top\beta_0)\,\partial_z\left(H_{d,\beta_0}((-\infty,t] \mid \cdot)\right)(X_i^\top\beta_0). \quad \text{(X.3)}$$

One could note that $\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0)\right] = 0$.

2. From the definitions we obtain

$$\begin{aligned}
&\widehat{U}_i(t,d;\beta_0) - \mathbb{E}\left[\widehat{U}_i(t,d;\beta_0) \mid Y_i,\delta_i,X_i\right] \\
&= \mathbf{1}\{Y_i \le t, \delta_i = d\}\frac{1}{n}\sum_{k=1}^n \left\{g^{-1}L((X_i - X_k)^\top\beta_0/g) - \mathbb{E}\left[g^{-1}L((X_i - X_k)^\top\beta_0/g) \mid X_i\right]\right\} \\
&\qquad + \frac{1}{n}\sum_{k=1}^n \Big\{\mathbf{1}\{Y_k \le t, \delta_k = d\}g^{-1}L((X_i - X_k)^\top\beta_0/g) \\
&\qquad\qquad - \mathbb{E}\left[\mathbf{1}\{Y_k \le t, \delta_k = d\}g^{-1}L((X_i - X_k)^\top\beta_0/g) \mid X_i\right]\Big\}.
\end{aligned}$$

The uniform rate follows by applying twice the Maximal Inequality (X.2) with $m = 1$ and α close to 1. Let us detail one these situations, the other one is similar. For any $w \in \mathbb{R}^p$, $t \ge 0$ and $d \in \{0,1\}$, let

$$f = \mathbf{1}\{\cdot \le t, \cdot = d\}L((w - \cdot)^\top\beta_0/g) - \mathbb{E}\left[\mathbf{1}\{\cdot \le t, \cdot = d\}L((w - \cdot)^\top\beta_0/g)\right]$$

a function of the variables Y, δ, X that depends on w, t, d. By the imposed assumptions, the family of such functions f indexed by w, t, d is Euclidean for a bounded envelope. Moreover, since for any real numbers a and b, $(a-b)^2 \le 2(a^2 + b^2)$, and L is bounded,

$$f^2 \le 2\{L((w - \cdot)^\top\beta_0/g) + \mathbb{E}\left[L((w - \cdot)^\top\beta_0/g)\right]\}.$$

One can apply inequality (X.2) with $m = 1$. It remains to bound the expectation on the right hand side. For this, up to a constant, one can use the bound

$$\mathbb{E}\left[L((X_1 - X_2)^\top\beta_0/g)\right] \lesssim g$$

and Jensen's inequality to deduce that the right hand side of the inequality (X.2) is of rate $g^{\alpha/2}$. The uniform rate of $V_{U,i}$ follows immediately. Modulo minor modifications and using also the assumption (IX.2) and Cauchy-Schwarz inequality, the uniform rate of $V_{\Delta,i}$ could be obtained by the same arguments. □

Lemma X.2. *Under the condition of Proposition 4.1:*

1.

$$\sup_{t,d} \Bigg\| \frac{1}{n(n-1)} \sum_{i \neq j} \mathbb{E}[\nabla_\beta U_i(t,d;\beta_0) \mid X_i] \mathbb{E}[\nabla_\beta U_j(t,d;\beta_0) \mid X_j]^\top \omega_{ij} - \mathbb{E}\left[\mathbb{E}[\nabla_\beta U_1(t,d;\beta_0) \mid X_1] \mathbb{E}[\nabla_\beta U_2(t,d;\beta_0) \mid X_2]^\top \omega_{12}\right] \Bigg\| = o_{\mathbb{P}}(1).$$

2.

$$\sup_{t,d} \left\| \frac{1}{n(n-1)} \sum_{i \neq j} \mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right] \{U_j + V_{U,j}\} \omega_{ij} \right\| = O_{\mathbb{P}}\left(n^{-1/2}\right).$$

3.

$$\sup_{t,d} \left\| \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} V_{\nabla,i} \{U_j + V_{U,j}\} \omega_{ij} \right\| = o_{\mathbb{P}}\left(n^{-1/2}\right).$$

Proof of Lemma X.2. 1. Use Hoeffding decomposition in degenerate $U-$statistics. Next use twice the maximal inequality (X.2) or Corollary 4 of Sherman (1994).

2. The assumptions guarantee the VC-class property for the class of functions

$$z \mapsto \partial_z \left(H_{d,\beta_0}((-\infty,t] \mid \cdot)\right)(z), \qquad t \in (-\infty,\infty),$$

for $d=0$ and $d=1$. The VC-class property for the class $\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right] U_j \, \omega_{ij}$, indexed by t and d, follows. Since

$$\mathbb{E}\left[\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right] U_j \, \omega_{ij}\right] = \mathbb{E}\left[\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right] \mathbb{E}[U_j(t,d;\beta_0) \mid X_j] \, \omega_{ij}\right] = 0,$$

Corollary 4 of Sherman (1994) provides the uniform rate $O_{\mathbb{P}}\left(n^{-1/2}\right)$ for the $U-$process

$$\mathcal{V}_n(t,d;\beta_0) = \frac{1}{n(n-1)} \sum_{i \neq j} \mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right] U_j(t,d;\beta_0)\, \omega_{ij}.$$

For the remaining part in the $U-$process investigated at this point, simplifying the notation $\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right]$ to $\mathbb{E}\left[\nabla_\beta U_i \mid X_i\right]$, we can write

$$\begin{aligned}
\frac{1}{n(n-1)} \sum_{i \neq j} \mathbb{E}\left[\nabla_\beta U_i \mid X_i\right] V_{U,j}\, \omega_{ij} &= \frac{g^{-1}}{n(n-1)} \sum_{i \neq j \neq k} \mathbb{E}\left[\nabla_\beta U_i \mid X_i\right] \\
&\times \left\{ \mathbf{1}\{Y_j \leq t, \delta_j = d\} [L((X_j - X_k)^\top \beta_0 / g) - \mathbb{E}(L((X_j - X)^\top \beta_0 / g) \mid X_j)] \right\} \omega_{ij} \\
&- \frac{g^{-1}}{n(n-1)} \sum_{i \neq j \neq k} \mathbb{E}\left[\nabla_\beta U_i \mid X_i\right] \left\{ [\mathbf{1}\{Y_k \leq t, \delta_k = d\} L((X_j - X_k)^\top \beta_0 / g) \right. \\
&\left. - \mathbb{E}(\mathbf{1}\{Y \leq t, \delta = d\} L((X_j - X)^\top \beta_0 / g) \mid X_j)] \right\} \omega_{ij} \\
&+ \text{diagonal terms of smaller order.} \\
&= g^{-1} [\mathcal{U}_{1n}(t,d;\beta_0) - \mathcal{U}_{2n}(t,d;\beta_0)] + \text{diagonal terms of smaller order.}
\end{aligned}$$

Next, we have to apply the Hoeffding decomposition to the third order $U-$processes $\mathcal{U}_{1n}(t,d;\beta_0)$ and $\mathcal{U}_{2n}(t,d;\beta_0)$ indexed by t and d. By the Maximal Inequality (X.2), the uniform rate of the degenerate third order $U-$processes in the Hoeffding decomposition of $\mathcal{U}_{1n}(t,d,\beta_0)$ and $\mathcal{U}_{2n}(t,d;\beta_0)$ is $O_{\mathbb{P}}(g^{\alpha/2} n^{-3/2})$. This rate, divided by g, is negligible compared to $O_{\mathbb{P}}(n^{-1/2})$. Now, let us denote $\upsilon_{1\,|\,i,j,k}$ and $\upsilon_{2\,|\,i,j,k}$ the functions defining $\mathcal{U}_{1n}(t,d;\beta_0)$ and $\mathcal{U}_{2n}(t,d;\beta_0)$. Moreover, for any $i \neq j$, let us denote

$$\mathbb{E}_{ij}[\cdots] = \mathbb{E}[\cdots \mid Y_i, \delta_i, X_i, Y_j, \delta_j, X_j] \quad \text{and} \quad \mathbb{E}_i[\cdots] = \mathbb{E}[\cdots \mid Y_i, \delta_i, X_i].$$

By construction, for each t and d,

$$\mathbb{E}_{ij}[\upsilon_{1\,|\,i,j,k}(t,d;\beta_0)] = \mathbb{E}_{ij}[\upsilon_{2\,|\,i,j,k}(t,d;\beta_0)] = 0. \tag{X.4}$$

Next, by the maximal inequality of Sherman (1994), the two remaining degenerate second order $U-$processes in the Hoeffding decomposition of $\mathcal{U}_{1n}(t,d;\beta_0)$ and $\mathcal{U}_{2n}(t,d;\beta_0)$ are of uniform rate is $O_{\mathbb{P}}(g^{\alpha/2} n^{-1})$. This rate, divided by g, is still neg-

ligible compared to $O_{\mathbb{P}}(n^{-1/2})$. Finally, the property (X.4) implies

$$\mathbb{E}_i[\upsilon_{l\,|\,i,j,k}(t,d;\beta_0)] = \mathbb{E}_j[\upsilon_{l\,|\,i,j,k}(t,d;\beta_0)] = 0, \quad \text{for} \quad l=1 \quad \text{and} \quad l=2.$$

Thus it remains to investigate the empirical processes defined by $\mathbb{E}_k[\upsilon_{1\,|\,i,j,k}(t,d;\beta_0)]$ and $\mathbb{E}_k[\upsilon_{2\,|\,i,j,k}(t,d;\beta_0)]$. By elementary calculations, uniformly with respect to t and d,

$$\begin{aligned}
\mathbb{E}_k[\upsilon_{1\,|\,i,j,k}(t,d;\beta_0)] &= \mathbb{E}_k\left[\mathbb{E}[\nabla_\beta U_i\,|\,X_i] H_{d,\beta_0}((-\infty,t]\,|\,X_j^\top\beta_0) L((X_j - X_k)^\top\beta_0/g)\,\omega_{ij}\right] \\
&- \mathbb{E}\left[\mathbb{E}[\nabla_\beta U_i \mid X_i] H_{d,\beta_0}((-\infty,t] \mid X_j^\top\beta_0)\mathbb{E}[L((X_j - X)^\top\beta_0/g) \mid X_j]\,\omega_{ij}\right] \\
&= g\mathbb{E}\left[\mathbb{E}[\nabla_\beta U_i \mid X_i]\omega_{ik} \mid X_k^\top\beta_0\right] H_{d,\beta_0}((-\infty,t] \mid X_k^\top\beta_0) f_{\beta_0}(X_k^\top\beta_0) \\
&- g\mathbb{E}\left\{\mathbb{E}\left[\mathbb{E}[\nabla_\beta U_i \mid X_i]\omega_{ij} \mid X_j^\top\beta_0\right] H_{d,\beta_0}((-\infty,t] \mid X_j^\top\beta_0) f_{\beta_0}(X_j^\top\beta_0)\right\} + O_{\mathbb{P}}(g^3).
\end{aligned}$$

On the other hand, uniformly with respect to t and d,

$$\begin{aligned}
\mathbb{E}_k[\upsilon_{2\,|\,i,j,k}(t,d;\beta_0)] &= \mathbb{E}_k\left[\mathbb{E}[\nabla_\beta U_i\,|\,X_i] H_{d,\beta_0}((-\infty,t]\,|\,X_k^\top\beta_0) L((X_j - X_k)^\top\beta_0/g)\,\omega_{ij}\right] \\
&- \mathbb{E}\left[\mathbb{E}[\nabla_\beta U_i \mid X_i]\mathbb{E}[H_{d,\beta_0}((-\infty,t] \mid X^\top\beta_0) L((X_j - X)^\top\beta_0/g) \mid X_j]\,\omega_{ij}\right] \\
&= g\mathbb{E}\left[\mathbb{E}[\nabla_\beta U_i \mid X_i]\omega_{ik} \mid X_k^\top\beta_0\right] H_{d,\beta_0}((-\infty,t] \mid X_k^\top\beta_0) f_{\beta_0}(X_k^\top\beta_0) \\
&- g\mathbb{E}\left\{\mathbb{E}\left[\mathbb{E}[\nabla_\beta U_i \mid X_i]\omega_{ij} \mid X_j^\top\beta_0\right] H_{d,\beta_0}((-\infty,t] \mid X_j^\top\beta_0) f_{\beta_0}(X_j^\top\beta_0)\right\} + O_{\mathbb{P}}(g^3).
\end{aligned}$$

As a consequence, uniformly with respect to t and d,

$$g^{-1}\left\{\mathbb{E}_k[\upsilon_{1\,|\,i,j,k}(t,d;\beta_0)] - \mathbb{E}_k[\upsilon_{2\,|\,i,j,k}(t,d;\beta_0)]\right\} = O_{\mathbb{P}}(g^2) = o_{\mathbb{P}}(n^{-1/2}).$$

Deduce that

$$\sup_{t,d}\left\|\frac{1}{n(n-1)}\sum_{i\neq j}\mathbb{E}\left[\nabla_\beta U_i(t,d;\beta_0) \mid X_i\right] V_{U,j}\,\omega_{ij}\right\| = o_{\mathbb{P}}\left(n^{-1/2}\right). \tag{X.5}$$

3. The same type of arguments as those used to derive the uniform rate (X.5) apply. The details are omitted. □

Let us introduce some more notation:

$$B_{\nabla^2,i} = \mathbb{E}\left[\nabla^2_{\beta\beta}\widehat{U}_i \mid X_i\right] - \mathbb{E}\left[\nabla^2_{\beta\beta}U_i \mid X_i\right]$$

and

$$V_{\nabla^2,i} = \nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta_0) - \mathbb{E}\left[\nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta_0) \mid X_i\right].$$

Lemma X.3. *Under the condition of Proposition 4.1:*

1.

$$\sup_i \sup_{t,d} |B_{\nabla^2,i}| = O_{\mathbb{P}}(g^2);$$

2. For any $\alpha \in (0,1)$,

$$\sup_i \sup_{t,d} \left\{ g^3 \left\| V_{\nabla^2,i} \right\| \right\} = O_{\mathbb{P}}(n^{-1/2} g^{\alpha/2});$$

3.

$$\sup_{t,d} \left| \frac{1}{n^2} \sum_{1\le i,j\le n} V_{\nabla^2,i}(t,d;\beta_0) U_j(t,d;\beta_0)\omega_{ij} \right| = o_{\mathbb{P}}(1);$$

4.

$$\sup_{t,d} \left| \frac{1}{n^2} \sum_{1\le i,j\le n} \mathbb{E}\left[\nabla^2_{\beta\beta}U_i(t,d;\beta_0) \mid X_i\right] U_j(t,d;\beta_0)\omega_{ij} \right| = o_{\mathbb{P}}(1).$$

Proof. 1. For a vector u, let $u^{\otimes 2}$ be the matrix $uu^\top$. For the uniform rate on the bias $B_{\nabla^2,i}$ let us write

$$\begin{aligned}
\mathbb{E}\left[\nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta_0) \mid X_i\right] &= H_{d,\beta_0}((-\infty,t] \mid X_i^\top\beta_0) \\
&\quad \times \mathbb{E}\left[\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X}_k \mid X_k^\top\beta_0]\right)^{\otimes 2} g^{-3} L''((X_i - X_k)^\top\beta_0/g) \,|\, X_i\right] \\
&\quad - \mathbb{E}\left[H_{d,\beta_0}((-\infty,t] \mid X_k^\top\beta_0)\left(\widetilde{X}_i - \mathbb{E}[\widetilde{X}_k \mid X_k^\top\beta_0]\right)^{\otimes 2}\right. \\
&\qquad \left. \times\, g^{-3} L''((X_i - X_k)^\top\beta_0/g) \mid X_i\right].
\end{aligned}$$

The uniform rate for $B_{\nabla^2,i}$ follows by integration by parts and Taylor expansion.

2. Let us simplify the notation and write $\mathcal{I}_i(t,d) = \mathbf{1}\{Y_i \le t; \delta_i = d\}$. To derive

the uniform rate of $g^3 \|V_{\nabla^2,i}\|$ it suffices to consider the following sum of empirical processes

$$
\begin{aligned}
g^3 \left[\nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta_0) - \mathbb{E}\left[\nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta_0) \mid X_i\right]\right] \\
= \frac{1}{n}\sum_{k=1}^{n} \Big\{\mathcal{I}_i(t,d)(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2} L''((X_i - X_k)^\top \beta_0/g) \\
- H_{d,\beta_0}((-\infty,t] \mid X_i^\top\beta_0)\mathbb{E}\left[(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2} L''((X_i - X_k)^\top \beta_0/g) \mid X_i\right]\Big\} \\
+ \frac{1}{n}\sum_{k=1}^{n} \Big\{\mathcal{I}_k(t,d)(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2} L''((X_i - X_k)^\top \beta_0/g) \\
- \mathbb{E}\left[H_{d,\beta_0}((-\infty,t] \mid X_k^\top\beta_0)(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2} L''((X_i - X_k)^\top \beta_0/g) \mid X_i\right]\Big\}
\end{aligned}
$$

and to apply the Maximal Inequality and a change of variables.

3. Up to uniformly negligible diagonal terms,

$$
\frac{g^3}{n^2}\sum_{1\le i,j\le n} V_{\nabla^2,i}(t,d;\beta_0)U_j(t,d;\beta_0)\omega_{ij}
$$

is equal to the following sum of two centered $U-$processes of order three

$$
\begin{aligned}
\frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k} \Big\{\mathcal{I}_i(t,d)(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2} L''((X_i - X_k)^\top \beta_0/g) \\
- H_{d,\beta_0}((-\infty,t] \mid X_i^\top\beta_0)\mathbb{E}\left[(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2} L''((X_i - X_k)^\top \beta_0/g) \mid X_i\right]\Big\} U_j\omega_{ij} \\
+ \frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k} \Big\{\mathcal{I}_k(t,d)(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2} L''((X_i - X_k)^\top \beta_0/g) \\
- \mathbb{E}\left[H_{d,\beta_0}((-\infty,t] \mid X_k^\top\beta_0)(\widetilde{X}_i - \widetilde{X}_k)^{\otimes 2} L''((X_i - X_k)^\top \beta_0/g) \mid X_i\right]\Big\} U_j\omega_{ij}
\end{aligned}
$$

for which we consider the Hoeffding decomposition. It is easy to check that the sum of the two $U-$processes of order 1 in the decomposition vanishes. Moreover, among the conditional expectations given a couple of variables, only the conditional expectations given the couples of observations (k,j) do not vanish. By the Maximal Inequality and standard calculations, the sum of degenerate $U-$processes of order 3 (resp. order 2) in the decomposition has the uniform rate $O_{\mathbb{P}}(n^{-3/2}g^{\alpha/2})$ (resp. $O_{\mathbb{P}}(n^{-1}g^{\alpha/2})$), where $\alpha \in (0,1)$. Since α could be close to 1 and $ng^3 \to \infty$, the stated result follows.

4. The same arguments as used for the previous rate apply. □

Let us recall the decomposition

$$\begin{aligned}\frac{1}{2}\nabla^2_{\beta\beta}\widehat{I}_n(t,d;\beta) &= \frac{1}{n^2}\sum_{1\le i,j\le n}\nabla_\beta\widehat{U}_i(t,d;\beta)\nabla_\beta\widehat{U}_j(t,d;\beta)^\top\omega_{ij}\\ &+ \frac{1}{n^2}\sum_{1\le i,j\le n}\nabla^2_{\beta\beta}\widehat{U}_i(t,d;\beta)\widehat{U}_j(t,d;\beta)\omega_{ij}\\ &= E_{n1}(t,d;\beta)+E_{n2}(t,d;\beta).\end{aligned}$$

Let (b_n) be a sequence tending to zero such that

$$\mathbb{P}[\|\widehat{\beta}-\beta_0\|\le b_n]\to 1.$$

Lemma X.4. *Under the condition of Proposition 4.1,*

$$\sup_{t,d}\sup_{\|\beta-\beta_0\|\le b_n}|E_{n1}(t,d;\beta)-E_{n1}(t,d;\beta_0)|=o_{\mathbb{P}}(1)$$

and

$$\sup_{t,d}\sup_{\|\beta-\beta_0\|\le b_n}|E_{n2}(t,d;\beta)-E_{n2}(t,d;\beta_0)|=o_{\mathbb{P}}(1).$$

Proof. Let us simplify the notation and write $\mathcal{I}_i=\mathbf{1}\{Y_i\le t;\delta_i=d\}$ and $H_i(\beta)=\mathbb{E}[\mathbf{1}\{Y_i\le t;\delta_i=d\}\mid X_i^\top\beta]$. Let us decompose

$$E_{n1}(t,d;\beta)=E_{n11}(t,d;\beta)+E_{n12}(t,d;\beta),$$

where the (r,q) entry of the matrix $E_{n11}(t,d;\beta)$ is

$$E_{n11}(t,d;\beta)_{rq}=\frac{g^{-4}}{(n)_4}\sum_{i\ne j\ne l\ne k}(\mathcal{I}_i-\mathcal{I}_k)(\mathcal{I}_j-\mathcal{I}_l)(\widetilde{X}_i-\widetilde{X}_k)_r(\widetilde{X}_j-\widetilde{X}_l)_qL'_{ik}(\beta,g)L'_{jl}(\beta,g)\omega_{ij},$$

and $E_{n1}(t,d;\beta)-E_{n11}(t,d;\beta)$. Here, $(n)_4=n(n-1)(n-2)(n-3)$, $L'_{ik}(\beta,g)=L'((X_i-X_k)^\top\beta/g)$, L' is the derivative of L, and for a column vector V, $(V)_r$ denotes its rth component. We apply the Hoeffding decomposition to the fourth order $U-$processes $g^4\,[E_{n11}(t,d;\beta)_{rq}-E_{n11}(t,d;\beta_0)_{rq}]$. The degenerate $U-$processes

of order four and three have the respective uniform rates $O_{\mathbb{P}}(n^{-2})$ and $O_{\mathbb{P}}(n^{-3/2})$. (These rates could be improved by a factor $g^{\alpha/2}$, but this is unnecessary at this stage.) Since $ng^3 \to \infty$, these degenerate $U-$processes will be negligible. It remains to study the degenerate $U-$processes of order 2, 1 and the mean and bound their variations when β gets close to β_0. Let us remark that, since ω_{ij} does not depend on β and given some symmetry arguments, conditioning with the observations (i,j) is similar to do it with the variables (k,l), (i,l) is similar to (j,k) and (i,k) to (j,l). That means, only three types of conditional expectations given two observations, as involved in the degenerate $U-$process of order two, have to be investigated. Let us denote

$$\mathbb{E}_{ij}[\cdots] = \mathbb{E}[\cdots \mid Y_i, \delta_i, X_i, Y_j, \delta_j, X_j] \quad \text{and} \quad \mathbb{E}_i[\cdots] = \mathbb{E}[\cdots \mid Y_i, \delta_i, X_i]$$

and, for any given components r and q, simply denote

$$e_{ikjl}(\beta) = e_{ikjl}(\beta)_{rq} = (\mathcal{I}_i - \mathcal{I}_k)(\mathcal{I}_j - \mathcal{I}_l)(\widetilde{X}_i - \widetilde{X}_k)_r(\widetilde{X}_j - \widetilde{X}_l)_q L'_{ik}(\beta, g)L'_{jl}(\beta, g)\omega_{ij}.$$

Then, by elementary properties of the conditional expectation,

$$\begin{aligned}
\mathbb{E}_{ij}[e_{ikjl}(\beta)] &= (\widetilde{X}_i)_r(\widetilde{X}_j)_q\omega_{ij}\mathbb{E}_i[(\mathcal{I}_i - H_k(\beta))L'_{ik}(\beta, g)]\mathbb{E}_j[(\mathcal{I}_j - H_l(\beta))L'_{jl}(\beta, g)] \\
&\quad - (\widetilde{X}_i)_r\omega_{ij}\mathbb{E}_i[(\mathcal{I}_i - H_k(\beta))L'_{ik}(\beta, g)]\mathbb{E}_j[\mathbb{E}[(\widetilde{X}_l)_q \mid X_l^\top\beta](\mathcal{I}_j - H_l(\beta))L'_{jl}(\beta, g)] \\
&\quad - (\widetilde{X}_j)_q\omega_{ij}\mathbb{E}_i[\mathbb{E}[(\widetilde{X}_k)_r \mid X_k^\top\beta](\mathcal{I}_i - H_k(\beta))L'_{ik}(\beta, g)]\mathbb{E}_j[(H_j(\beta) - H_l(\beta))L'_{jl}(\beta, g)] \\
&\quad + \omega_{ij}\mathbb{E}_i[\mathbb{E}[(\widetilde{X}_k)_r \mid X_k^\top\beta](\mathcal{I}_i - H_k(\beta))L'_{ik}(\beta, g)]\mathbb{E}_j[\mathbb{E}[(\widetilde{X}_l)_q \mid X_l^\top\beta](\mathcal{I}_j - H_l(\beta))L'_{jl}(\beta, g)].
\end{aligned}$$

For each of the eight conditional expectations $\mathbb{E}_i$ or $\mathbb{E}_j$ on the right hand side of the last equation we could apply twice Lemma X.5 to bound their variation when β gets close to β_0. For instance,

$$\begin{aligned}
&\mathbb{E}_i[\mathbb{E}[(\widetilde{X}_k)_r \mid X_k^\top\beta](\mathcal{I}_i - H_k(\beta))g^{-2}L'_{ik}(\beta, g)] \\
&\qquad = \mathbb{E}_i[\mathbb{E}[(\widetilde{X}_k)_r \mid X_k^\top\beta_0](\mathcal{I}_i - H_k(\beta_0))g^{-2}L'_{ik}(\beta_0, g)] + R_i,
\end{aligned}$$

where,

$$\sup_{1\leq r\leq p-1}\sup_{t,d}\sup_{\|\beta-\beta_0\|\leq b_n}|R_i| = \{1+\|X_i\|\}o_{\mathbb{P}}(1),$$

and the $o_{\mathbb{P}}(1)$ factor does not depend on X_i. Moreover, the conditional expectation $\mathbb{E}_i[\mathbb{E}[(\widetilde{X_k})_r \mid X_k^\top\beta_0](\mathcal{I}_i - H_k(\beta_0))g^{-2}L'_{ik}(\beta_0,g)]$ is uniformly bounded. Recall that, by our assumptions, $\sup_{1\leq i\leq n}\|\widetilde{X_i}\| = O_{\mathbb{P}}(\log n)$. Deduce that

$$\sup_{t,d}\sup_{\|\beta-\beta_0\|\leq b_n}|\mathbb{E}_{ij}[e_{ikjl}(\beta)]-\mathbb{E}_{ij}[e_{ikjl}(\beta_0)]|^2 = \{1+\|X_i\|^2+\|X_j\|^2\}g^8 o_{\mathbb{P}}(1) = g^8\log^2 n\; o_{\mathbb{P}}(1), \tag{X.6}$$

and the $o_{\mathbb{P}}(1)$ factor does not depend on X_i or X_j. Moreover, the conditional expectation $g^{-4}\mathbb{E}_{ij}[e_{ikjl}(\beta)]/\log n$ is uniformly bounded. Thus the conditional expectations $g^{-4}\mathbb{E}_i[e_{ikjl}(\beta)]/\log n$, $g^{-4}\mathbb{E}_j[e_{ikjl}(\beta)]/\log n$ and the expectation $g^{-2}\mathbb{E}[e_{ikjl}(\beta)]/\log n$ are also uniformly bounded. Recall that $g^a\log n\to 0$ for any $a>0$. Gathering facts, the contribution of $\mathbb{E}_{ij}[e_{ikjl}(\beta)] - \mathbb{E}_{ij}[e_{ikjl}(\beta_0)]$ to the right hand side of the maximal inequality, applied with the degenerate $U-$process of order two in the Hoeffding decomposition of $g^4\left[E_{n11}(t,d;\beta)_{rq} - E_{n11}(t,d;\beta_0)_{rq}\right]$, is of rate $O(g^{4\alpha}\log^\alpha n)$ with $\alpha\in(0,1)$ that could be arbitrarily close to 1. Finally, this will produce a contribution of uniform rate $O_{\mathbb{P}}(n^{-1}g^{4\alpha-4}\log^\alpha n) = o_{\mathbb{P}}(1)$ for the $U-$processes of order two.

Next, let us investigate the variations of $\mathbb{E}_{il}[e_{ikjl}(\beta)]$. In this case, one still decompose $\mathbb{E}_{il}[e_{ikjl}(\beta)]$ in a sum of terms, each of them under the form of products involving conditional expectations given only one observation. Then, the same uniform rate as in equation (X.6) could be derived. Moreover, by similar arguments, the conditional expectation $g^{-4}\mathbb{E}_l[e_{ikjl}(\beta)]/\log n$ and the expectation $g^{-4}\mathbb{E}[e_{ikjl}(\beta)]/\log n$ could be shown to be uniformly bounded. One could conclude that $\mathbb{E}_{il}[e_{ikjl}(\beta)] - \mathbb{E}_{il}[e_{ikjl}(\beta_0)]$ will behave similarly to $\mathbb{E}_{ij}[e_{ikjl}(\beta)] - \mathbb{E}_{ij}[e_{ikjl}(\beta_0)]$ uniformly in a neighborhood of β_0.

To complete the analysis of the degenerate $U-$processes of order 2, it remains

to investigate the variations of $\mathbb{E}_{ik}[e_{ikjl}(\beta)]$. We have

$$
\begin{aligned}
&g^{-2}\mathbb{E}_{ik}[e_{ikjl}(\beta)_{rq}] = (\mathcal{I}_i - \mathcal{I}_k)(\widetilde{X}_i - \widetilde{X}_k)_r L'_{ik}(\beta, g)\mathbb{E}_i[(\mathcal{I}_j - \mathcal{I}_l)(\widetilde{X}_j - \widetilde{X}_l)_q g^{-2} L'_{jl}(\beta, g)\omega_{ij}] \\
&= (\mathcal{I}_i - \mathcal{I}_k)(\widetilde{X}_i - \widetilde{X}_k)_r L'_{ik}(\beta, g)\mathbb{E}_i[\mathbb{E}[(\widetilde{X}_j)_q \mathcal{I}_j \omega_{ij} \mid X_i, X_j^\top \beta]\mathbb{E}[g^{-2} L'_{jl}(\beta, g) \mid X_i, X_j^\top \beta]] \\
&- (\mathcal{I}_i - \mathcal{I}_k)(\widetilde{X}_i - \widetilde{X}_k)_r L'_{ik}(\beta, g)\mathbb{E}_i[\mathbb{E}[\mathcal{I}_j \omega_{ij} \mid X_i, X_j^\top \beta]\mathbb{E}_l[(\widetilde{X}_l)_q \mid X_i, X_l^\top \beta] g^{-2} L'_{jl}(\beta, g)] \\
&- (\mathcal{I}_i - \mathcal{I}_k)(\widetilde{X}_i - \widetilde{X}_k)_r L'_{ik}(\beta, g)\mathbb{E}_i[\mathbb{E}[\mathcal{I}_l \mid X_i, X_l^\top \beta]\mathbb{E}_l[(\widetilde{X}_j)_q \omega_{ij} \mid X_i, X_j^\top \beta] g^{-2} L'_{jl}(\beta, g)] \\
&+ (\mathcal{I}_i - \mathcal{I}_k)(\widetilde{X}_i - \widetilde{X}_k)_r L'_{ik}(\beta, g)\mathbb{E}_i[\mathbb{E}[(\widetilde{X}_l)_q \mathcal{I}_l \omega_{ij} \mid X_i, X_l^\top \beta]\mathbb{E}[g^{-2} L'_{jl}(\beta, g) \mid X_i, X_l^\top \beta]].
\end{aligned}
$$

Using the fact that $\sup_{1 \le i \le n} \|X_i\| = O_{\mathbb{P}}(\log n)$, by Lemma X.5,

$$
\mathbb{E}[g^{-2} L'_{jl}(\beta, g) \mid X_i, X_j^\top \beta] = \mathbb{E}[g^{-2} L'_{jl}(\beta_0, g) \mid X_j^\top \beta_0] + o_{\mathbb{P}}(\log n),
$$

uniformly with respect to $\|\beta - \beta_0\| \le b_n$. On the other hand, using arguments as in Lemma X.5, for some suitable functions γ_β,

$$
\mathbb{E}[\gamma_\beta(X_i^\top \beta, X_k^\top \beta; t, d) g^{-2}[L'_{ik}(\beta, g)]^2] = \mathbb{E}[\gamma_\beta(X_i^\top \beta_0, X_k^\top \beta_0; t, d) g^{-2}[L'_{ik}(\beta, g)]^2] + o_{\mathbb{P}}(\log n),
$$

uniformly. Moreover, for uniformly bounded functions γ_β, since $L'(\cdot)$ is bounded and integrable, $\mathbb{E}[\gamma_\beta(X_i^\top \beta_0, X_k^\top \beta_0; t, d) g^{-2}[L'_{ik}(\beta, g)]^2]$ is uniformly bounded. Repeating the arguments several times, deduce that

$$
\mathbb{E}\left[\sup_{t,d} \sup_{\|\beta - \beta_0\| \le b_n} \{\mathbb{E}_{ik}[e_{ikjl}(\beta)_{rq}] - \mathbb{E}_{ik}[e_{ikjl}(\beta_0)_{rq}]\}^2\right] = g^6 O(\log^4 n),
$$

and thus, by the Maximal Inequality, $\mathbb{E}_{ik}[e_{ikjl}(\beta)] - \mathbb{E}_{ik}[e_{ikjl}(\beta_0)]$ is uniformly negligible in a neighborhood of β_0. The degenerate $U-$processes of order 1 and the mean could be handled with similar arguments. For the mean, one slight difference is that instead of taking the supremum of quantities like $\|X_i\|$ that appears in the bound given by Lemma X.5, one has to integrate them out. This avoids to pay the price of a $\log n$ factor in the bounds and hence yields $\mathbb{E}[e_{ikjl}(\beta)] - \mathbb{E}[e_{ikjl}(\beta_0)] = o_{\mathbb{P}}(1)$ uniformly with respect to t and d, and uniformly over $o_{\mathbb{P}}(1)$ neighborhoods of β_0.

Up to a factor that converges to 1, the quantity $E_{n12}(t, d; \beta)$ contains the diagonal

terms of $E_{n1}(t,d;\beta)$ and thus is negligible. The arguments for $E_{n2}(t,d;\beta)$ are very similar and hence we omit the details. Let us only mention that for the degenerate $U-$process of order 1 and for the mean, we use the fact that, by our assumptions and for suitable functions γ_β, quantities like $\mathbb{E}_i\left[\gamma_\beta(X_i^\top\beta, X_k^\top\beta; t, d)g^{-3}L''_{ik}(\beta, g)\right]$ are uniformly bounded. □

Lemma X.5. *Let $L^{(\kappa)}(\cdot)$ denote the κth derivative of $L(\cdot)$, $\kappa \in \{0,1,2\}$. Let b_n, $n \geq 1$ be a positive sequence of real numbers converging to zero. Let $z \mapsto \lambda_\beta(z;t,d)$, $z \in \mathbb{R}$ be a class of κ times differentiable functions indexed by β, t and d such that*

$$\sup_{\|\beta-\beta_0\|\leq b_n}\sup_{t,d}\sup_{z\in\mathbb{R}}\left|\partial_z^{(\kappa)}[\lambda_\beta(\cdot;t,d)f_\beta(\cdot)](z) - \partial_z^{(\kappa)}[\lambda_{\beta_0}(\cdot;t,d)f_{\beta_0}(\cdot)](z)\right| = 0,$$

and $\partial_z^{(\kappa)}[\lambda_{\beta_0}(\cdot;t,d)f_{\beta_0}(\cdot)](\cdot)$ is uniformly bounded, where $\partial_z^{(\kappa)}$ denotes the derivative of order κ with respect to z for $\kappa \in \{0,1,2\}$. If the function $z \mapsto \partial_z^{(\kappa)}[\lambda_{\beta_0}(\cdot;t,d)f_{\beta_0}(\cdot)](z)$ is Lipschitz continuous with constant C independent of t and d, then

$$\sup_{t,d}\sup_{\|\beta-\beta_0\|\leq b_n}\Big|\mathbb{E}\left[\lambda_\beta(X^\top\beta;t,d)g^{-\kappa-1}L^{(\kappa)}((x-X)^\top\beta/g)\right] - \mathbb{E}\left[\lambda_{\beta_0}(X^\top\beta_0;t,d)g^{-\kappa-1}L^{(\kappa)}((x-X)^\top\beta_0/g)\right]\Big| = \{1+\|x\|\}o_{\mathbb{P}}(1), \quad \text{(X.7)}$$

and the $o_{\mathbb{P}}(1)$ factor does not depend on x.

Proof of Lemma X.5. Let simplify notation and write $\lambda_\beta(\cdot)$ instead of $\lambda_\beta(\cdot;t,d)$. First we consider the case $\kappa = 0$. By a standard change of variables and the stated assumptions, uniformly with respect to t, d, $\|\beta-\beta_0\| \leq b_n$, for any $x \in \mathcal{X}$,

$$\begin{aligned}
\mathbb{E}\left[\lambda_\beta(X^\top\beta)g^{-1}L((x-X)^\top\beta/g)\right] &= \int_{\mathbb{R}}\lambda_\beta(z)g^{-1}L((x^\top\beta-z)/g)f_\beta(z)dz \\
&= \int_{\mathbb{R}}(\lambda_\beta f_\beta)(x^\top\beta-uh)L(u)du \\
&= \int_{\mathbb{R}}(\lambda_{\beta_0} f_{\beta_0})(x^\top\beta-uh)L(u)du + o(1) \\
&= \int_{\mathbb{R}}(\lambda_{\beta_0} f_{\beta_0})(x^\top\beta_0-uh)L(u)du + O(b_n) + \|x\|O(b_n) \\
&= \mathbb{E}\left[\lambda_{\beta_0}(X^\top\beta_0)g^{-1}L((x-X)^\top\beta_0/g)\right] + \{1+\|x\|\}o(1).
\end{aligned}$$

When $\kappa = 1$, before applying the uniform continuity condition (X.7) and the Lipschitz property, we use integration by parts to deduce

$$\int_{\mathbb{R}} \lambda_\beta(z) f_\beta(z) g^{-2} L'((x^\top\beta - z)/g)dz = \int_{\mathbb{R}} \partial_z[\lambda_\beta f_\beta](z) g^{-1} L((x^\top\beta - z)/g)dz.$$

Next we can continue as in the case $\kappa = 0$. The case $\kappa = 2$ is similar and hence we omit the details. □

Lemma X.6. *Under the Assumptions of Proposition 5.1, for any $b_n = O_{\mathbb{P}}(n^{-1/2})$,*

$$\sup_{\|\beta-\beta_0\|\le b_n} \sup_{t\le\tau} \sup_{x\in\overline{\mathcal{X}}} |\Delta_{1n}(t,x;\beta) - \Delta_{1n}(t,x;\beta_0)| = o_{\mathbb{P}}(n^{-1/2}h^{-1/2})$$

and

$$\sup_{\|\beta-\beta_0\|\le b_n} \sup_{t\le\tau} \sup_{x\in\overline{\mathcal{X}}} |\Delta_{2n}(t,x;\beta,\beta_0)| = o_{\mathbb{P}}(n^{-1/2}h^{-1/2}).$$

Proof of Lemma X.6. First, we investigate the uniform rate of Δ_{2n}. Let

$$\xi(Y,\delta;t) = \mathbf{1}\{Y \le t\}\delta \quad \text{or} \quad \xi(Y,\delta;t) = \mathbf{1}\{Y \ge t\}, \qquad t \le \tau,$$

and let $\widetilde{H}_\beta(t \mid z)$ be any of $H_{1,\beta}((-\infty,t] \mid z) = \mathbb{E}\left[\mathbf{1}\{Y \le t\}\delta \mid X^\top\beta = z\right]$ or

$$H_\beta([t,\infty) \mid z) = \mathbb{E}\left[\mathbf{1}\{Y \ge t\} \mid X^\top\beta = z\right], \qquad t \le \tau.$$

By standard changes of variables,

$$\begin{aligned}
&\mathbb{E}\left[h^{-1}\xi(Y,\delta;t)\left\{K((x-X)^\top\beta/h) - K((x-X)^\top\beta_0/h)\right\}\right] \\
&\quad = \mathbb{E}\left[h^{-1}\widetilde{H}_\beta(t \mid X^\top\beta)K((x-X)^\top\beta/h) - \widetilde{H}_{\beta_0}(t \mid X^\top\beta_0)K((x-X)^\top\beta_0/h)\right] \\
&= \int_{\mathbb{R}} K(u)\left[\widetilde{H}_\beta(t \mid x^\top\beta - uh) f_\beta(x^\top\beta - uh) - \widetilde{H}_{\beta_0}(t \mid x^\top\beta - uh) f_{\beta_0}(x^\top\beta - uh)\right] du \\
&+ \int_{\mathbb{R}} K(u)\left[\widetilde{H}_{\beta_0}(t \mid x^\top\beta - uh) f_{\beta_0}(x^\top\beta - uh) - \widetilde{H}_{\beta_0}(t \mid x^\top\beta_0 - uh) f_{\beta_0}(x^\top\beta_0 - uh)\right] du.
\end{aligned}$$

Since $\|\beta - \beta_0\| \leq b_n$, by the Lipschitz property of $\widetilde{H}_{\beta_0}(t \mid \cdot) f_{\beta_0}(\cdot)$, one could deduce

$$\sup_{\|\beta-\beta_0\|\leq b_n} \sup_{t\leq\tau} \sup_{x\in\overline{\mathcal{X}}} \left| \mathbb{E}\left[h^{-1}\xi(Y,\delta;t)\left\{ K((x-X)^\top\beta/h) - K((x-X)^\top\beta_0/h) \right\} \right] \right| \lesssim b_n.$$

For the uniform rate for Δ_{1n}, let it suffice to consider the empirical process defined by the family of functions

$$\mathcal{F}_n = \{\xi(\cdot,\cdot;t)\left[K((\cdot - x)^\top\beta/h) - K((\cdot - x)^\top\beta_0/h) \right] : t \leq \tau, x \in \overline{\mathcal{X}}, \|\beta - \beta_0\| \leq b_n\}.$$

By the result of Nolan & Pollard (1987), Pakes and Pollard (1989), see also Sherman (1994) and van der Vaart & Wellner (1996), this class is a VC-class for a constant envelope, with constants A and V independent of n. By the Maximal Inequality applied for some α close to 1, the fact that $K(\cdot)$ is bounded, the uniform bound on the expectation of the function from $\mathcal{F}_n$ derived above, one could deduce that

$$\sup_{\|\beta-\beta_0\|\leq b_n} \sup_{t\leq\tau} \sup_{x\in\overline{\mathcal{X}}} |\Delta_{1n}(t,x;\beta) - \Delta_{1n}(t,x;\beta_0)| = O_{\mathbb{P}}(n^{-1/2}h^{-1}b_n^{\alpha/2}) = o_{\mathbb{P}}(n^{-1/2}h^{-1/2}).$$

□

Lemma X.7. *Under the condition of Proposition 4.1*

$$\int \mathcal{V}_{1n}(t,d;\beta_0) dF_{n,Y,\delta}(t,d) = \int \mathcal{V}_{1n}(t,d;\beta_0) dF_{Y,\delta}(t,d) + o_{\mathbb{P}}(n^{-1/2}),$$

Proof. Let

$$v_{1j}(t,d;\beta_0) = \mathbb{E}\left\{ \mathbb{E}\left[\nabla_\beta U(t,d;\beta_0) \mid X \right] \omega(X - X_j) \mid X_j \right\} U_j(t,d;\beta_0)$$

so that

$$\mathcal{V}_{1n}(t,d;\beta_0) = \frac{1}{n} \sum_{1\leq j\leq n} v_{1j}(t,d;\beta_0).$$

Then,

$$\int \mathcal{V}_{1n}(t,d;\beta_0)dF_{n,Y,\delta}(t,d) = \frac{n-1}{n}\frac{1}{n(n-1)}\sum_{1\le i\ne j\le n} v_{1j}(Y_i,\delta_i;\beta_0) + \frac{1}{n^2}\sum_{1\le j\le n} v_{1j}(Y_j,\delta_j;\beta_0)$$
$$= \frac{n-1}{n}V_{n11} + V_{n12}.$$

By the law of large numbers, it is easy to see that $V_{n12} = O_{\mathbb{P}}(n^{-1})$. On the other hand,

$$\int v_{1j}(t,d;\beta_0)dF_{Y,\delta}(t,d)$$
$$= \mathbb{E}\left[\mathbb{E}\left\{\mathbb{E}\left[\nabla_\beta U(\tilde{Y},\tilde{\delta};\beta_0) \mid X\right]\omega(X-X_j) \mid X_j,\tilde{Y},\tilde{\delta}\right\}U_j(\tilde{Y},\tilde{\delta};\beta_0) \mid Y_j,\delta_j,X_j\right].$$

Next note that, since

$$\mathbb{E}\left[v_{1j}(Y_i,\delta_i;\beta_0) \mid Y_i,\delta_i,X_i\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left\{\mathbb{E}\left[\nabla_\beta U(\mid Y_i,\delta_i;\beta_0) \mid X,Y_i,\delta_i\right]\omega(X-X_j) \mid X_j\right\}\mathbb{E}[U_j(t,d;\beta_0) \mid X_j] \mid Y_i,\delta_i,X_i\right]$$
$$= 0$$

and

$$\mathbb{E}\left[\int v_{1j}(t,d;\beta_0)dF_{Y,\delta}(t,d) \mid Y_i,\delta_i,X_i\right] = 0,$$

we have

$$\mathbb{E}\left[v_{1j}(Y_i,\delta_i;\beta_0) - \int v_{1j}(t,d;\beta_0)dF_{Y,\delta}(t,d) \mid Y_i,\delta_i,X_i\right] = 0$$

Moreover, by construction,

$$\mathbb{E}\left[v_{1j}(Y_i,\delta_i;\beta_0) - \int v_{1j}(t,d;\beta_0)dF_{Y,\delta}(t,d) \mid Y_j,\delta_j,X_j\right] = 0.$$

This means that $V_{n11} - \int \mathcal{V}_{1n}(t,d;\beta_0)dF_{Y,\delta}(t,d)$ is a degenerate $U-$statistics of order 2 and thus, by a simple variance calculation,

$$V_{n11} = \int \mathcal{V}_{1n}(t,d;\beta_0)dF_{Y,\delta}(t,d) + O_{\mathbb{P}}(n^{-1}).$$

Gathering the rates, the statement follows. □

Additional References

1. Nolan, D., and Pollard, D. (1987). $U-$processes : Rates of convergence. *Annals of Statistics* **15**, 780–799.

2. Pakes, A., and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027–1057.

3. van der Vaart, A.D. & Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New-York.

ANNEXE 2 (Modèle dernière page de thèse)

VU :

Le Directeur de Thèse
(Nom et Prénom)

PATILEA Valentin

VU :

Le Responsable de l'École Doctorale

LION Jean-Marie

VU pour autorisation de soutenance

Rennes, le

Le Président de l'Université de Rennes 1

Guy CATHELINEAU

VU après soutenance pour autorisation de publication :

Le Président de Jury,
(Nom et Prénom)

Résumé

Dans cette thèse, nous étudions des modèles définis par des équations de moments conditionnels. Une grande partie de modèles statistiques (régressions, régressions quantiles, modèles de transformations, modèles à variables instrumentales, *etc.*) peuvent se définir sous cette forme. Nous nous intéressons au cas des modèles avec un paramètre à estimer de dimension finie, ainsi qu'au cas des modèles semi paramétriques nécessitant l'estimation d'un paramètre de dimension finie et d'un paramètre de dimension infinie. Dans la classe des modèles semi paramétriques étudiés, nous nous concentrons sur les modèles à direction révélatrice unique qui réalisent un compromis entre une modélisation paramétrique simple et précise, mais trop rigide et donc exposée à une erreur de modèle, et l'estimation non paramétrique, très flexible mais souffrant du fléau de la dimension. En particulier, nous étudions ces modèles semi paramétriques en présence de censure aléatoire. Le fil conducteur de notre étude est un contraste sous la forme d'une $U-$statistique, qui permet d'estimer les paramètres inconnus dans des modèles généraux.

Mots clés. Analyse de Survie ; Direction révélatrice unique ; Données censurées ; Equations de moments conditionnels ; Fonctionnelles de Kaplan-Meier ; Lissage par noyau ; Méthodes itératives ; Modèles de régression ; Réduction de la dimension ; Rééchantillonnage ; Régression pénalisée ; $U-$statistiques

Abstract

In this dissertation we study statistical models defined by condition estimating equations. Many statistical models could be stated under this form (mean regression, quantile regression, transformation models, instrumental variable models, *etc.*). We consider models with finite dimensional unknown parameter, as well as semiparametric models involving an additional infinite dimensional parameter. In the latter case, we focus on single-index models that realize an appealing compromise between parametric specifications, simple and leading to accurate estimates, but too restrictive and likely misspecified, and the nonparametric approaches, flexible but suffering from the curse of dimensionality. In particular, we study the single-index models in the presence of random censoring. The guiding line of our study is a $U-$statistics which allows to estimate the unknown parameters in a wide spectrum of models.

Keywords. Bootstrap ; Censoring ; Conditional moment equations ; Dimension reduction ; Iterative methods ; Kaplan-Meier functionals ; Kernel smoothing ; Penalized regression ; Regression models ; Single-index assumptions ; Survival analysis ; $U-$statistics