Nonparametric inference for the mixture cure model when the cure status is partially known

Wende Clarence Safari

PhD Thesis

2022

University of A Coruña

UNIVERSIDADE DA CORUÑA

Nonparametric inference for the mixture cure model when the cure status is partially known

Wende Clarence Safari

PhD Thesis

2022

Supervisors:
María Amalia Jácome Pumar
Ignacio López de Ullibarri

Doctoral Program:
Statistics and Operational Research
University of A Coruña

UNIVERSIDADE DA CORUÑA

UNIVERSIDADE DA CORUÑA

The undersigned, María Amalia Jácome Pumar and Ignacio López de Ullibarri, certify that they are the advisors of the Doctoral Thesis entitled “Nonparametric inference for the mixture cure model when the cure status is partially known”, developed by Wende Clarence Safari at the University of A Coruña (Department of Mathematics), as part of the interuniversity PhD program (UDC, USC and UVigo) of Statistics and Operations Research, and hereby give their consent to the author to proceed with the thesis presentation and the subsequent defense.

Los abajo firmantes, María Amalia Jácome Pumar e Ignacio López de Ullibarri, hacen constar que son los directores de la Tesis Doctoral titulada “Nonparametric inference for the mixture cure model when the cure status is partially known”, realizada por Wende Clarence Safari en la Universidade da Coruña (Departamento de Matemáticas) en el marco del programa interuniversitario (UDC, USC y UVigo) de doctorado en Estadística e Investigación Operativa, dando su consentimiento para que la autora proceda a su presentación y posterior defensa.

Os abaixo asinantes, María Amalia Jácome Pumar e Ignacio López de Ullibarri, fan constar que son os directores da Tese de Doutoramento titulada “Nonparametric inference for the mixture cure model when the cure status is partially known”, realizada por Wende Clarence Safari na Universidade da Coruña (Departamento de Matemáticas) no marco do programa interuniversitario (UDC, USC e UVigo) de doutoramento en Estatística e Investigación de Operacións, dando o seu consentimento para que a autora proceda á sua presentación e posterior defensa.

A Coruña, 01 de junio de 2022

Advisor: Prof. María Amalia Jácome **Advisor:** Prof. Ignacio López de Ullibarri

PhD student: Wende Clarence Safari

UNIVERSIDADE DA CORUÑA

The public defense of the Doctoral Thesis entitled "Nonparametric inference for the mixture cure model when the cure status is partially known", developed by Wende Clarence Safari and supervised by Prof. María Amalia Jácome Pumar and Prof. Ignacio López de Ullibarri, will be held on 30th June, 2022, at the Faculty of Computer Sciences at the University of A Coruña, with the examining committee:
President: Prof. Paul Yingwei Peng (Queen's University, Canada)
Board Member: Prof. Eni Musta (Universiteit van Amsterdam, Netherlands)
Secretary: Prof. Ricardo José Cao Abad (Universidade da Coruña)

A Coruña, 30th June 2022

PhD committee:

Prof. Paul Yingwei Peng Prof. Eni Musta Prof. Ricardo José Cao Abad

Advisors:

Prof. María Amalia Jácome Prof. Ignacio López de Ullibarri

PhD student:

Wende Clarence Safari

I dedicate this thesis to my mother, Grace Lazaro Mwipopo, who has been my constant source of inspiration. She has given me the drive and discipline to handle any task with aspiration and determination.

"When you learn, teach, when you get, give."
Maya Angelou

"A good head and good heart are always a formidable combination. But when you add to that a literate tongue or pen, then you have something very special."
Nelson Mandela

Acknowledgement

First of all, I am very grateful to my supervisors for their valuable advice, their continued support, and patience during my PhD journey. Their vast knowledge and experience have encouraged me throughout my PhD research and personal life.

Professor María Amalia Jácome Pumar has gone beyond the call of thesis advisor to the role of academic mother, making efforts not only to explain the concepts of cure models but also to introduce me to prominent researchers in the field. Her charisma, perseverance and faith were crucial to birth of this thesis and to my upbringing as an independent researcher of the future.

I am very grateful to Professor Ignacio López de Ullibarri for his rentless efforts to guide me through research and thesis writing. His immeasurable expertise, suggestions and critiques have played a significant role in completing this research work.

I offer my sincere thanks to Professor Ricardo Cao for giving me the opportunity to pursue this PhD. Also, I appreciate his technical support during various phases of this PhD. Many thanks also go to Professor Jacobo de Uña and Professor María Carmen Suarez Cadarso, as their discussions and comments during the predefense have greatly improved this thesis. I would like to thank all members in the MODES group. To name but a few, Ana López Cheda, Ana Almécija Pereda, Cristina Muiños Roca and Luis Fernando Rodríguez. It is their kind help and support that have made my study and life in Spain a wonderful time.

I would like to acknowledge Professor Anouar El Ghouch for his warmest welcome and valuable suggestions during my research stay at the ISBA, Université Catholique de Louvain, in Belgium.

I am extremely grateful to my parents for their prayers, caring, and continuing

support. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study.

I am very thankful to my younger sister, Merciana, for keeping me informed about family issues in her own special way. I also thank my younger brother, Bongracious, for his thoughtfulness and for providing me with information on ongoing activities at home. I am truly blessed to have Nelson Ochieng, da Angel Kidulile, Jacqueline Materu (Adol) and John Mapilya as my closest people all the time. I also thank my extended family and friends for their support throughout my academic trajectory.

I do not want to miss this opportunity to thank Manolo, Hugo and Gabi for being so kind and loving. I am really grateful to have such a family in Spain.

Special thanks are extended to Jim Todd, Leacky Kamau, Innocent Mboya, Neema Mosha, Zakile Mfumbilwa, and the whole team of the International Biometric Society (IBS) of Tanzania for their encouragement and enthusiasm during my PhD.

I would like to thank God the Almighty for His guidance throughout my PhD journey. Also for granting me countless blessings and knowledge. I will keep on trusting Him for my future. Mungu awabariki ninyi nyote!

Institutional acknowledgement

This research has been partially supported by the Spanish Ministerio de Economía y Competitividad (MINECO) under Grants MTM2014-52876-R and MTM2017-82724-R, by the Spanish Ministerio de Ciencia e Innovación (MICINN) under Grant PID2020-113578RB-I00, by the Agencia Galega de Innovación (GAIN) under Grant IN845D 2020/26, by the Xunta de Galicia through Centro Singular de Investigación de Galicia accreditation under Grants ED431G 2019/01 and through the Grupos de Referencia Competitiva under Grants ED431C 2016-015 and ED431C2020-014 and in part by the European Union (European Regional Development Fund-ERDF).

The author particularly thanks the contracts financed by the research group MODES from August, 2018 to September, 2021, and by the CITIC from October, 2021 to March, 2022.

The work has been partially carried out at the Université Catholique de Louvain (UCLouvain) through INDITEX-UDC 2020 grant. The author acknowledges accommodation support from the Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA) of the UCLouvain during three months stay at the Institute as a visiting researcher.

The author is grateful to Dr. Juan Ruiz Bañobre and Dr. Angel Díaz-Lagares from the Cancer Epigenomics of Translational Medical Oncology (OMT) group, Health Research Institute of Santiago (IDIS) and the University Hospital of Santiago (CHUS), Spain, for providing the sarcoma and breast cancer datasets. Also, many thanks to Andrés Paz-Ares Rodríguez (General Director of Public Health, Xunta de Galicia), Xurxo Hervada Vidal (General Deputy Director of Information on Health and Epidemiology, Xunta de Galicia) and Benigno Rosón Calvo (General Deputy director of the SERGAS information system) for providing the COVID-19 data.

Abstract

Classical analysis of time-to-event data assumes that all individuals will eventually experience the event of interest. However, when there is evidence of long-term survivors, cure models should be used instead. They assume that the population of individuals is made up of two distinct groups: those who will and those who will not experience the event. A common assumption in cure models is that there is no additional information about the cure status, and the cure indicator is modelled as a latent variable. But this is not entirely valid in many cases, when some censored individuals can be identified as cured, for example, based on a diagnostic test or if the observed lifetime is larger than a cure threshold. Mixture cure models have been usually estimated using parametric or semiparametric models. Recently, a completely nonparametric approach was introduced under the classical assumption that the cure status in unknown. This PhD thesis proposes a novel extension of nonparametric mixture cure models to incorporate the additional information about the cure status. Suitable nonparametric estimators for the main functions are proposed, together with a rough procedure for checking the validity of the model.

Resumen

Los métodos clásicos de análisis de tiempos de vida asumen que todos los individuos experimentarán el suceso de interés. Sin embargo, cuando hay evidencia de la presencia de supervivientes a largo plazo o curados, se deberán usar en su lugar los modelos de curación. Estos asumen que la población de individuos se puede dividir en dos grupos: los que experimentarán el suceso y los que no lo harán. Cuando se aplican los modelos de curación se asume que no se dispone de información adicional sobre el estado de cura, y el indicador de cura se modeliza en consecuencia como una variable latente. Sin embargo esto no es necesariamente cierto en muchos casos, en los que algunos individuos censurados se pueden identificar como curados, basándose por ejemplo en un test diagnóstico o si el tiempo de vida supera un determinado umbral. Los modelos de curación de tipo mixtura se han estimado normalmente usando técnicas paramétricas o semiparamétricas. Recientemente se ha propuesto un enfoque completamente no paramétrico para los modelos de curación de tipo mixtura, bajo la hipótesis clásica de que se desconoce completamente si un sujeto está curado. Esta tesis propone una extensión a los modelos no paramétricos de curación de tipo mixtura, en la que se incorporaría la información adicional disponible sobre el estado de cura. Se proponen estimadores no paramétricos de las principales funciones, así como un sencillo método para comprobar la validez del modelo.

Resumo

Os métodos clásicos de análisis de tempos de vida asumen que todos os individuos experimentarán o suceso de interese. Con todo, cando hai evidencia da presenza de supervivientes a longo prazo ou curados, deberansen usar no seu lugar os modelos de curación. Estes asumen que a poboación de individuos pódese dividir en dous grupos: os que experimentarán o suceso e os que non. Cando se aplican os modelos de curación, asúmese que para os individuos censurados non se dispón de información adicional sobre o estado de cura, e o indicador de cura se modeliza en consecuencia como unha variable latente. Mais isto non é necesariamente certo en moitos casos, nos que algúns individuos censurados pódense identificar como curados, baseándose por exemplo nun test diagnóstico ou se o tempo de vida supera un determinado valor. Os modelos de curación de tipo mixtura estimáronse normalmente usando técnicas paramétricas ou semiparamétricas. Recentemente propúxose un enfoque completamente non paramétrico para os modelos de curación de tipo mixtura, baixo a hipótese clásica de que se descoñece completamente se un individuo está curado. Esta tese propón unha extensión aos modelos non paramétricos de curación de tipo mixtura, na que se incorporará información adicional dispoñible sobre o estado de cura. Propóñense estimadores non paramétricos das principais funcións, así como un posible procedemento para avaliar a validez do modelo.

Preface

The objective of this thesis is to propose a completely nonparametric methodology in the mixture cure model when the cure status is partially known. A summary of each chapter is provided below.

Chapter 1 begins with a brief overview of classical survival analysis, also introduces the concepts of censoring and truncation. A thorough review of cure model is provided with two estimation methods exist in the literature explained in details – namely mixture cure models and non-mixture cure models. Estimation of key quantities of interest under the mixture cure model (survival function, cure probability and latency function) are introduced. Three real datasets are carefully described in the last section of the chapter. These data have distinct features and are analyzed thoroughly in subsequent chapters.

Chapter 2, the main contribution of this thesis, introduces a novel nonparametric estimator of the conditional survival function in the MCM for right censored data when the cure status is partially known. The estimator is developed for a setting with a single continuous covariate but can be extended to multiple covariates. It extends the estimator by Beran (1981), which ignores cure status information. An almost sure representation is obtained, from which the strong consistency and asymptotic normality of the estimator are derived. Asymptotic expressions for the bias and variance of the proposed estimator demonstrate a reduction in the variance with respect to Beran's estimator. A simulation study shows that the proposed estimator performs better than others for an ample range of covariate values, if the bandwidth parameter is suitably chosen. A bootstrap bandwidth selector is proposed. Finally, the estimator is applied to a real dataset studying survival of sarcoma patients. This work was published in Safari et al. (2021).

Chapter 3 covers the second contribution of the thesis. A new estimator of the probability of the cure is proposed. The proposed estimator extends that of Xu and Peng (2014) to the context in which some censored individuals can be considered as cured. The estimator is shown to be strongly consistent and asymptotically normally distributed. Two alternative estimators are also presented, which have not been previously studied in the literature. The first of them derives from the competing risks approach. The MCM with cure partially known can be considered as a special competing risks model in which there are two types of competing events, the event of interest and the cure. The main idea underlying the second alternative estimator is that, since the cure indicator is a binary variable, the probability of cure can be written also as the conditional mean of the cure indicator. As a consequence, nonparametric regression methods can be applied to estimate this conditional mean. However, under right random censorship, the indicator of cure is not known for all individuals, since for many censored individuals it is not known if they will finally experience the event of interests or not. The application of regression methods in this context requires handling missing data in the response variable (cure indicator). Simulations are performed to evaluate the finite sample performance of all these estimators, and apply them to the analysis of two datasets related to, respectively, survival of breast cancer patients and length of hospital stay of COVID-19 patients requiring intensive care. This work has been submitted for publication (Safari et al., 2022a).

The problem of estimating the latency function in the MCM when the cure status information is partially available is addressed in Chapter 4. A latency estimator that extends the nonparametric estimator studied in López-Cheda et al. (2017b) to the case when the cure status is partially available is proposed. The asymptotic properties of the proposed estimator are established, and its performance is studied via a simulation study. Finally, the estimator is applied to a medical dataset to study the length of hospital stay of COVID-19 patients requiring intensive care. This work has been submitted for publication (Safari et al., 2022b).

An important feature of the proposed estimators is that they are consistent only under the assumption of conditional independence of the survival and censoring times. In Chapter 5, a simple nonparametric procedure is proposed to assess how plausible the independence assumption is when the censoring rate is high. It relies on the fact that the difference between the proposed MCM-based kernel estimator of the cure probability and the regression-based estimator should be large under independence. The motivation comes from the fact that under independence assumption, only the proposed MCM-based estimator provides a good approximation of the conditional cure probability. Meanwhile, for the regression-based estimator to be consistent, the independence assumption must not be fulfilled.

To close the thesis, Chapter 6 provides some general conclusions and identifies issues that deserve further research. The proofs of theoretical results in Chapters 2 – 4 are relegated to Appendices A – C, respectively.

The results in Chapter 2 have been published in Safari et al. (2021). The results of each of Chapters 3 and 4 have gathered in manuscripts submitted for publication in international peer-reviewed journals, as listed below.

1. Safari, W. C., López-de-Ullibarri, I., and Jácome, M. A. (2021). A product-limit estimator of the conditional survival function when cure status is partially known. *Biometrical Journal*, **63**(5):984 – 1005.

2. Safari, W. C., López-de-Ullibarri, I., and Jácome, M. A. (2022a). Nonparametric kernel estimation of the probability of cure in a mixture cure model when the cure status is partially observed. *First revision in Statistical Methods in Medical Research.*

3. Safari, W. C., López-de-Ullibarri, I., and Jácome, M. A (2022b). Latency function estimation under the mixture cure model when the cure status is available. *Submitted.*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Survival analysis refers to the analysis of the length of time until the occurrence of a well-defined event of interest (survival time). Others refer to such techniques as time-to-event or event history analysis. Since data can only be collected over a specified time period, it is likely that not all subjects will have the event by the end of study period. So the actual survival times for some subjects are unknown. This phenomenon is referred to as censoring. The use of standard statistical methods in these data might generate results that have some level of bias because important information would be left out. Survival data structure can be viewed as consisting of two important measures such as survival times and censoring status.

Censoring is divided into three general types: right censoring, left censoring, and interval censoring (Klein and Moeschberger, 2003; Kalbfleisch and Prentice, 2011). The most common type of censoring encountered in survival analysis data is right censoring. Right censoring arises when the event of interest is not observed within the study duration, and it would happen later than the observed time. This may occur, for example, if an individual drops out of a study before the event of interest happens, because of the end of study or loss to follow-up. Left censoring occurs when the event of interest has occurred prior to the observation time, but it is unknown exactly when. Interval censoring occurs when the actual event times are unknown, as the event is only known to be located between two known time points.

The survival time can also be subject to right and/or left truncation (Klein and Moeschberger, 2003). Left truncation occurs when data is only recorded for individuals whose survival time exceeds a random time (i.e., left truncation time).

Right truncation occurs when data is only recorded for individuals whose survival time proceeds a random time (i.e., right truncation time). When both left and right truncation are present, this is known as double truncation. This thesis is focused on situations where the event times are subject to right censoring, as this is the most frequent scenario encountered in the literature.

In classical survival analysis, it is commonly assumed that the event of interest will always happen if there is a sufficient follow-up time. However, there are many examples where the event will not occur for all individuals. For instance, some cancer patients will never relapse or die from cancer, some bank customers will never default in loan repayment, etc. Those whose event is certain not to occur are considered "statistically cured" (or *long-term survivors*) and those who will experience the event are known to be "uncured" (or *susceptible*) subjects. Cure models (Legrand, 2021; Peng and Yu, 2021) have been developed to address this issue. There are two types of cure models: mixture cure model (MCM) and non-mixture cure model (NMCM). Both mixture and non-mixture cure models aim to estimate the survival function of the population.

The MCM, initially proposed by Boag (1949), has received much attention in recent years. It assumes that the population is a mixture of cured and susceptible individuals. Note that here a "cured" individual is defined as being free of experiencing the event of interest, not necessarily cured in medical terms. The goal is to model the probability of cure and the survival function of the uncured subjects, also called latency. Extensive research has been conducted for the standard MCM from either a (semi)parametric point of view (Maller and Zhou, 1992; Amico and Van Keilegom, 2018; Patilea and Van Keilegom, 2020, among others), or from a completely nonparametric approach (Xu and Peng, 2014; López-Cheda et al., 2017a,b).

The NMCM, also known as promotion time cure model or bounded cumulative hazard model, has a proportional hazard model structure. It was firstly introduced by Yakovlev et al. (1993) and later discussed by (Chen et al., 1999; Chen et al., 2002) and Ibrahim et al. (2001), to name a few. Tsodikov et al. (2003) provided a review of existing methodology of statistical inference based on the NMCM. They highlighted two advantages of the NMCM: it presents a much more biologically meaningful interpretation of the results of the data analysis. Moreover, it is easy in computations due to its simple structure for the survival function which can provide a certain technical advantage when developing maximum likelihood estimation procedures.

Although MCM and NMCM represent two different modeling approaches and differences have been underlined in the literature, the two methods are related and have a meaningful connection (Legrand, 2021; Peng and Yu, 2021). Besides, there exists a number of works that have unified the MCM and NMCM. Interested readers are referred to Amico (2018) for an overview of the related research.

The absence of an individual's cure status (i.e., cured, uncured) is an important challenge for cure models. A subject whose event is observed is known to be uncured. However, censoring prevents from observing whether a censored subject would eventually experience the event or not. This hinders the classification of censored observations as cured or uncured. In this situation, it is customary to assume no additional information on the cure status of the censored individuals, thus, to model the cure status as a latent variable. Nonetheless, there are situations where some of the censored individuals can be identified to be immune to the event of interest, that is, to be cured. For example, diagnostic procedures in medical studies are available to provide further information on whether a subject suffering a curable illness can be considered as cured and therefore will not die from that disease. Also, for some types of cancer it is extremely unlikely to have any recurrence later than a given time after treatment, known as cure threshold, and consequently those patients with observed time surpassing the cure threshold can be considered relapse-free, therefore cured from the recurrence (Taylor, 1995). Another example of situations with individuals known to be cured from the event is the analysis of hospital bed and intensive care unit (ICU) occupancy (López-Cheda et al., 2021). In this, it is important to estimate the distribution of the time a patient will be in the hospital ward or ICU. Specifically, modeling the time a patient stays in the hospital ward until admitted to the ICU. In the language of cure models, all patients who have died or have been discharged from the hospital bed without entering the ICU are censored and are known to be cured from the ICU admission. Accurate estimates of the trajectory of patients and their length of stay from one hospital facility (ward) to another (ICU) are crucial for efficient resource management by healthcare authorities particularly during outbreaks of epidemic diseases such as the novel COVID-19 disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

In this kind of examples with a subgroup identified as cured from the event, there are three groups of observations: the event times (individuals who experienced the event during the follow-up time); the “regular” censored times (those who neither experienced the event by the end of the study nor classified as were cured); and a new third group, the “cured” censored times (of those acknowledged as cured from the event). Just modeling the data under the usual cure model framework will not take advantage of this additional cure status information given by the third group.

Several authors have explored cure models when the cure status is known for some censored observations. Under the MCM framework, Nieto-Baraja and Yin (2008) proposed a Bayesian semiparametric model for survival data with a cure fraction. They considered a fixed cure threshold, so that observations censored at times larger than it are assumed to correspond to cured subjects. A semiparametric approach based on a Cox proportional hazards cure model when cure information is partially known was studied by Wu et al. (2014). Bernhardt (2016) proposed a method where the probability of cure is estimated using a logistic model and the latency is estimated by a flexible accelerated failure time (AFT) model. Under the NMCM, Chen and Du (2018) proposed a method where both the cure probability and the latency function are estimated by nonparametric smoothing spline functions. Recently, Lin and Huang (2019) extended a Cox model to investigate the effects of covariates on the cure probability with different values of the cure threshold. All these authors have shown that ignoring the known cured information can lead to biased and inaccurate estimates.

The estimation of the MCM using a completely nonparametric estimator, in the spirit of the estimator by Kaplan and Meier (1958), has been only addressed by Laska and Meisner (1992) by defining cure as survival beyond a fixed cure threshold, and Betensky and Schoenfeld (2001) whose method allows for the observed cured censoring times to be random. Neither of these nonparametric proposals can handle covariates.

None of the aforementioned works considered a completely nonparametric MCM in the presence of known cure status and covariates. Therefore, the goal of this thesis is to develop a completely nonparametric kernel methodology in the context of MCM when the cure status is partially known. The remainder of this chapter is dedicated to the introduction of the concepts underlying the research presented in this work. Section 1.1 presents some basic concepts in the standard survival models and a review of nonparametric estimators of the survival function. Section 1.2 introduces the concept of cure in survival data and an overview of the nonparametric estimation and testing in the MCM. Different resampling plans for bootstraping in the presence of cured individuals are presented in Section 1.3. Since nonparametric kernel methods critically depend on a smoothing parameter (or bandwidth), its choice in the MCM is addressed in Section 1.4. In Section 1.5, the basic notation used in the MCM when the cure status is partially known is introduced. Additional notation will be introduced in Chapters 2 – 5, whenever necessary. Three datasets that will be used to illustrate the proposed methodologies are also described in Section 1.6.

1.1 Standard survival model

1.1.1 Model notation and the basic framework

This section starts by introducing the notation in the classical survival model. Let Y be the time until the event of interest and $\boldsymbol{X}$ a vector of covariates. The conditional cumulative distribution and survival functions of Y are defined as $F(t \mid \boldsymbol{x}) = P(Y \leq t \mid \boldsymbol{X} = \boldsymbol{x})$, and $S(t \mid \boldsymbol{x}) = 1 - F(t \mid \boldsymbol{x}) = P(Y > t \mid \boldsymbol{X} = \boldsymbol{x})$, respectively. Suppose the survival time Y is censored by a random variable C, and the conditional distribution of C is denoted by $G(t \mid \boldsymbol{x}) = P(C \leq t \mid \boldsymbol{X} = \boldsymbol{x})$. The random variables Y and C are assumed to be conditionally independent given $\boldsymbol{X} = \boldsymbol{x}$. In the presence of right censoring, only the pair (T, δ) is observed where $T = \min(Y, C)$ and $\delta = \mathbf{1}(Y < C)$. The conditional distribution of the observed time T is denoted by $H(t \mid \boldsymbol{x}) = P(T \leq t \mid \boldsymbol{X} = \boldsymbol{x})$. Without loss of generality, hereafter, let us consider a continuous covariate X with density function $m(x)$.

1.1.2 Kaplan-Meier estimator

In this subsection, the problem of estimating nonparametrically the survival function in a setting without covariates is revised. Kaplan and Meier (1958) were the first to propose a nonparametric estimator of the survival function, $S(t)$, which is defined by

$$\widehat{S}_n(t) = \prod_{i=1}^{n} \left(1 - \frac{\delta_{[i]} \mathbf{1} \left(T_{(i)} \leq t \right)}{n - i + 1} \right) \tag{1.1}$$

where $\delta_{[i]}, i = 1, \ldots, n$, are the concomitants of the ordered observed times $T_{(1)} \leq \cdots \leq T_{(n)}$. Kaplan and Meier (1958) showed that the estimator in equation (1.1) is the nonparametric maximum likelihood estimator of $S(t)$.

Basic properties of the Kaplan-Meier (KM) estimator, also known as the product-limit (PL) estimator, are the following:

1. The KM estimator is a right-continuous step function with jumps at the event times (i.e, $\delta_i = 1$). The magnitude of the jumps can be expressed as
$$\begin{aligned}\widehat{S}_n(T_{(i-1)}) - \widehat{S}_n(T_{(i)}) &= \frac{\delta_{[i]}}{n}\prod_{j=1}^{i-1}\left(1+\frac{1-\delta_{[j]}}{n-j}\right) = \frac{\delta_{[i]}}{n-i+1}\prod_{j=1}^{i-1}\left(1-\frac{\delta_{[j]}}{n-i+1}\right)\\ &= \frac{\delta_{[i]}}{n-i+1}\widehat{S}_n(T_{(i-1)}).\end{aligned}$$

2. When there are no censored data values, the KM estimator reduces to the empirical survival function:
$$\widetilde{S}_n(t) = \prod_{i=1}^{n}\frac{(n-i)\mathbf{1}\left(Y_{(i)} \le t\right)}{n-i+1} = 1 - \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(Y_i \le t).$$

3. The product of the KM estimators of the survival function of the variables Y and C coincides with the empirical distribution of the observed time T:
$$\begin{aligned}1-\widehat{H}_n(t) &= \left(1-\widehat{F}_n(t)\right)\left(1-\widehat{G}_n(t)\right)\\ &= \prod_{i=1}^{n}\left(1-\frac{\delta_{[i]}\mathbf{1}\left(T_{(i)} \le t\right)}{n-i+1}\right)\prod_{i=1}^{n}\left(1-\frac{\left(1-\delta_{[i]}\right)\mathbf{1}\left(T_{(i)} \le t\right)}{n-i+1}\right)\\ &= \prod_{i=1}^{n}\left(1-\frac{\mathbf{1}(T_{(i)} \le t)}{n-i+1}\right) = 1-\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(T_i \le t).\end{aligned}$$

The asymptotic properties of the KM estimator have been extensively studied in the literature. Kaplan and Meier (1958) and Efron (1967) showed that the KM estimator is a weakly consistent estimator of the survival function $S(t)$. Breslow and Crowley (1974) proved the convergence and the asymptotic normality and Földes and Rejto (1981) proved the uniform consistency.

1.1.3 Beran estimator

When covariates are available, one can consider the estimation of the conditional survival function using the estimator by Beran (1981), which represents a direct extension from the KM estimator to the regression context. The Beran estimator of $S(t \mid x)$ is given:
$$\widehat{S}_h(t \mid x) = \prod_{i=1}^{n}\left(1-\frac{\delta_{[i]}B_{h[i]}(x)\mathbf{1}\left(T_{(i)} \le t\right)}{\sum_{j=i}^{n}B_{h[j]}(x)}\right) \tag{1.2}$$

where $\delta_{[i]}, i = 1, \ldots, n$, are the concomitants of the ordered observed times $T_{(1)} \leq \cdots \leq T_{(n)}$. The weights $B_{h[i]}(x)$ are defined as:

$$B_{h[i]}(x) = \frac{K_h\left(x - X_{[i]}\right)}{\sum_{j=1}^{n} K_h(x - X_j)} \tag{1.3}$$

where $X_{[i]}$ is the concomitant of the ordered observed time $T_{(i)}, i = 1, \ldots, n$, and $K_h(\cdot) = K(\cdot/h)/h$ is a kernel function $K(\cdot)$ rescaled with bandwidth h. The bandwidth h tends to 0 as $n \to \infty$. In the following, alluding to the well-known Nadaraya-Watson (NW) kernel regression estimator (Nadaraya, 1964a; Watson, 1964), these weights will be referred to as the NW weights.

Beran estimator (1.2) inherits the same properties as its unconditional counterpart (1.1):

1. In the simplest case, when all the data are observed completely (no censoring), $\widehat{S}_h(t \mid x)$ reduces to the kernel estimator of the conditional survival function (Nadaraya, 1964b).

2. In case of no covariates, $\widehat{S}_h(t \mid x)$ reduces to the KM estimator (1.1).

3. The product of the Beran estimators for the survival functions of the variables Y and C is equal to the conditional empirical estimator of the observed time T:

$$\begin{aligned} 1 - \widehat{H}_h(t \mid x) &= (1 - \widehat{F}_h(t \mid x))(1 - \widehat{G}_h(t \mid x)) \\ &= \prod_{i=1}^{n}\left(1 - \frac{\delta_{[i]} B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x)}\right) \prod_{i=1}^{n}\left(1 - \frac{(1-\delta_{[i]}) B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x)}\right) \\ &= \prod_{i=1}^{n}\left(1 - \frac{B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x)}\right) = 1 - \sum_{i=1}^{n} B_{h[i]}(x) \mathbf{1}(T_{(i)} \leq t). \end{aligned}$$

Large sample properties of Beran estimator have been studied extensively (Beran, 1981; Dabrowska, 1987; Dabrowska, 1989; González-Manteiga and Cadarso-Suárez, 1994; Van Keilegom and Veraverbeke, 1997 Iglesias-Pérez and González-Manteiga, 1999).

1.2 Mixture cure model

1.2.1 Model notation and formulation

The standard survival model assumes that if there is no censoring, then at some point all individuals in the study will experience the event of interest. Hence, the survival function is a proper survival function:

$$\lim_{t \to \infty} S(t \mid x) = 0.$$

However, when survival data contain a cure fraction, two different types of observations are considered: those who experience the event and therefore known to be uncured, and those who will never experience the event and thus be considered as cured. When considering a cure fraction, it is assumed that the time-to-event for a cured subject is $Y = \infty$, in order to represent the fact that the event never will happen. Let $\nu = \mathbf{1}(Y = \infty)$ denote the indicator of being cured from the event. Note that ν is partially observed because $\delta = 1$ implies $\nu = 0$, but ν is usually unknown for the censored observations. As a consequence, when t becomes large, a fraction of the observations is still event-free and the survival function is improper:

$$\lim_{t \to \infty} S(t \mid x) > 0.$$

An informal way to identify the possible presence of cured individuals is to look at a plateau on the right tail of the survival curve of the KM estimator (Andrei and Asselain, 1996). If there is a clear long plateau in the right tail, showing the fact that the survival function is improper, and one can assume that (almost) all observations in the plateau correspond to cured observations. Figure 1.1 shows, for a simulated data example with a cure proportion, the KM curves estimated under the standard survival model and a cure model.

The MCM considers that the population of interest is actually a mixture between the cured and uncured subgroups. The probability of being cured is $1 - p(x) = P(Y = \infty \mid X = x)$, and the conditional survival function of the uncured individuals, also known as latency, is the continuous function $S_0(t \mid x) = P(Y > t \mid Y < \infty, X = x)$. The MCM writes the survival function $S\left(t \mid x\right)$ as

$$S\left(t \mid x\right) = 1 - p(x) + p(x) S_0\left(t \mid x\right). \tag{1.4}$$

Assuming model (1.4), the cure rate and the latency can be written in terms of the survival function as follows:

$$1 - p(x) = \lim_{t \to \infty} S(t \mid x) > 0, \quad S_0(t \mid x) = \frac{S(t \mid x) - (1 - p(x))}{p(x)}.$$

Therefore, the availability of a suitable estimator of $S(t \mid x)$ would yield appropriate estimators of the cure probability and the latency directly.

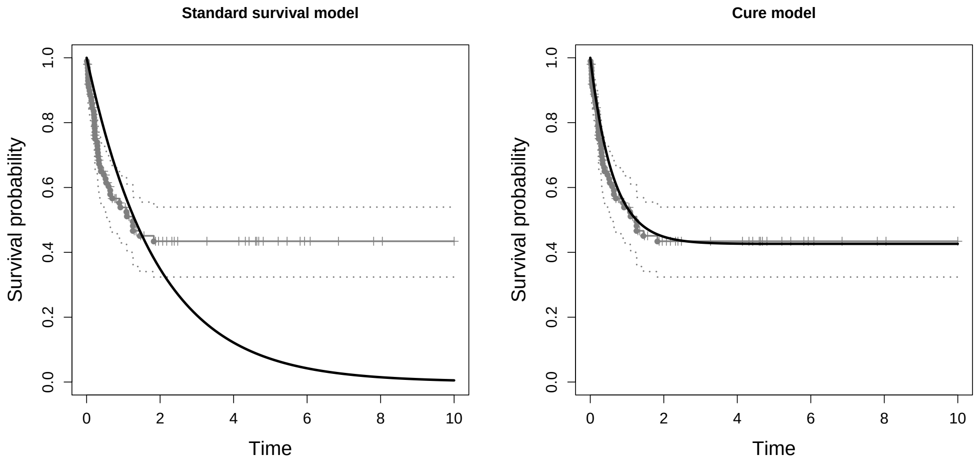

Figure 1.1: KM curves with 95% confidence intervals (gray lines) and true survival functions (black lines), estimated under the standard survival model (left) and using a cure model (right). Events are marked with a solid circle and censored observations with a cross.

1.2.2 Model identifiability

One key issue in cure models is identifiability. This arises because of the lack of cure status information at the end of the follow-up period, resulting in difficulties in distinguishing models with high incidence of susceptibles and long tails of the latency distribution from low incidence of susceptibles and short tails of the latency distribution (Li et al., 2001). Following the argumentation of Hanin and Huang (2014), who discussed in detail the identifiability of the MCM, model (1.4) is identifiable if the latency function is proper. Thus, it is assumed that $\lim_{t\to\infty} S_0(t \mid x) = 0$ for all x. This condition is similar to the zero-tail constraint in Taylor (1995), López-Cheda et al. (2017a) and other works.

1.2.3 Nonparametric estimation and inference in the classical mixture cure model

Maller and Zhou (1992) were the first to propose a completely nonparametric methodology of the MCM. However, their method does not handle covariates. Wang et al. (2012) proposed a MCM with a nonparametric form in the cure probability, and to ensure model identifiability, they assumed a nonparametric proportional hazards model for the hazard function. The estimation was carried out by an expectation-maximization (EM) algorithm for a penalized likelihood.

They defined the smoothing spline function estimates as the minimizers of the penalized likelihood. Patilea and Van Keilegom (2020) employed a nonparametric form for the latency but the cure probability was specified using a logistic regression model.

Xu and Peng (2014) extended the existing work of Maller and Zhou (1992) by proposing a nonparametric cure probability estimator in the presence of a continuous covariate. Their starting point was based on the Beran estimator in (1.2). They introduced the following kernel type cure probability estimator:

$$1 - \widehat{p}_h(x) = \widehat{S}_h(T^1_{(n)} \mid x) = \prod_{i=1}^{n} \left(1 - \frac{\delta_{[i]} B_{h[i]}(x)}{\sum_{j=i}^{n} B_{h[j]}(x)} \right) \tag{1.5}$$

where $T^1_{(n)}$ is the largest uncensored observed time. This estimator was also studied by López-Cheda et al. (2017a). The asymptotic properties (i.i.d representation, the strong consistency and convergence to a Gaussian process) of the estimator in (1.5) have been studied extensively by Xu and Peng (2014) and López-Cheda et al. (2017a).

As for the latency, López-Cheda et al. (2017b) proposed the following estimator:

$$\widehat{S}_{0,h}(t \mid x) = \frac{\widehat{S}_h(t \mid x) - (1 - \widehat{p}_h(x))}{\widehat{p}_h(x)} \tag{1.6}$$

where $\widehat{S}_h(t \mid x)$ is the generalized product-limit estimator of $S(t \mid x)$ in (1.2) and $1 - \widehat{p}_h(x)$ is the cure probability estimator in (1.5). In addition, López-Cheda et al. (2017b) derived the asymptotic properties of their estimator. Hereafter, the estimator in (1.5) will be referred to XP estimator while the estimator in (1.6) will be cited as the LC estimator.

Müller and Van Keilegom (2019) proposed a goodness-of-fit test based on the XP estimator as a diagnostic tool to investigate misspecification of the parametric model assumption for the cure probability in the MCM. López-Cheda et al. (2020) proposed a nonparametric covariate hypothesis test for the probability of cure in the MCM and applied a bootstrap method to approximate the null distribution of the test statistic.

Note that when the covariate X is a high-dimensional vector $\boldsymbol{X}$, the nonparametric kernel estimators are affected by the so-called “curse of dimensionality”. This phenomenon is essentially due to the sparsity of data as the dimension of the vector increases, which results to dramatically bad performances of the nonparametric kernel estimators. Therefore, other kind of models such as the single-index Cox MCM (Amico et al., 2019) have been considered in the literature. Single-index models are said to be much more flexible than purely parametric models also do not suffer from the curse of dimensionality. Concretely, the single-index model assumes that there exists an unknown link function that links Y with $\boldsymbol{X}$ by a single score (called *index*) mean regression model. Interested readers are referred to Amico et al. (2019) for more details.

1.3 Bootstrap procedures

Ever since its introduction by Efron (1981), bootstrap resampling procedures for censored data have become a widely used technique for confidence interval construction, testing hypotheses and data driven bandwidth selection. There are two equivalent resampling algorithms for bootstrapping: "simple method" and "obvious method". These methods were studied in detail by Reid (1981) and Akritas (1986), among others. Li and Datta (2001) extended Efron's resampling approaches to incorporate covariates.

In the presence of right censoring, covariates and a cured fraction, López-Cheda et al. (2017a,b), using similar ideas to those in Li and Datta (2001), suggested two equivalent bootstrap resampling algorithms under the MCM. They proposed a simple weighted bootstrap method (López-Cheda et al., 2017a), which is detailed as follows. Generate $\{X_1^*, \ldots, X_n^*\}$ from the empirical distribution of $\{X_1, \ldots, X_n\}$. Next, for each $i = 1, \ldots, n$, generate (T_i^*, δ_i^*) from the weighted empirical distribution $\widehat{F}_g\left(. \mid X_i^*\right)$ of $\{\left(T_1, \delta_1\right), \ldots, \left(T_n, \delta_n\right)\}$ given by

$$\widehat{F}_g\left(t, d \mid X_i^*\right) = \sum_{j=1}^{n} B_{gj}\left(X_i^*\right) \mathbf{1}\left(T_j \leq t, \delta_j \leq d\right),$$

where $B_{gj}(x)$ are the NW weights in (1.3) computed with the pilot bandwidth g. López-Cheda et al. (2017b) also suggested an obvious weighted bootstrap method. This procedure is carried out as follows. First, generate $\{X_1^*, \ldots, X_n^*\}$ from the empirical distribution of $\{X_1, \ldots, X_n\}$. Given X_i^*, for each $i = 1, \ldots, n$, generate Y_i^* from $\widehat{S}_{0,g}\left(t \mid X_i^*\right)$ with probability $\widehat{p}_g(X_i^*)$ and $Y_i^* = \infty$ otherwise. Generate C_i^* from $1 - \widehat{G}_g\left(t \mid X_i^*\right)$ for $i = 1, \ldots, n$. Here, $\widehat{S}_{0,g}\left(t \mid x\right)$ is the estimator of $S_0\left(t \mid x\right)$ defined in (1.6) and $\widehat{p}_g(x)$ is the XP estimator in (1.5), and $\widehat{G}_g\left(t \mid x\right)$ is the generalized product-limit estimator of the censoring distribution, all of them computed with the pilot bandwidth g. Finally, define $T_i^* = \min(Y_i^*, C_i^*)$ and $\delta_i^* = \mathbf{1}\left(Y_i^* < C_i^*\right)$, for $i = 1, \ldots, n$. So, the bootstrap sample is formed as $\{\left(X_i^*, T_i^*, \delta_i^*\right), i = 1, \ldots, n\}$.

López-Cheda et al. (2017a) assumed that $1 - G(t \mid x) = 1 - G(t)$ for all x, t, and consequently the KM estimator $1 - \widehat{G}_n(t)$ was considered for estimating the censoring distribution. Besides, López-Cheda et al. (2017a) found no relevant differences between generating $\{X_1^*, \ldots, X_n^*\}$ from the empirical distribution of $\{X_1, \ldots, X_n\}$ or fixing $X_i^* = X_i$, for $i = 1, \ldots, n$, so they considered the latter one for computational efficiency.

1.4 Bandwidth selection methods

One of important issue in nonparametric kernel estimation is the choice of an appropriate bandwidth, as it is well known that the performance of kernel estimators depends heavily on this parameter. A wide variety of methods for bandwidth selection of kernel estimators are available in the literature, cross-validation (CV), plug-in and bootstrap methods being the most common. In the context of MCM, the finite-sample behavior of the CV bandwidth selector for the estimator of the cure rate in equation (1.5) was found to be unsatisfactory, as it is highly variable and tends to undersmooth (López-Cheda et al., 2017a).

Basically, a plug-in bandwidth selector tries to estimate the dominant term of the mean integrated squared error (MISE) or mean squared error (MSE) of the kernel estimator, and obtain the optimal bandwidth that minimizes the estimated dominant term of the MISE or MSE. However, the optimal bandwidth h cannot be easily estimated in practice, due to the unknown, hard-to-estimate quantities found in the corresponding error criterion function. This tends to a process that seems to be more complicated than the original estimation problem.

Bootstrap procedures have been successfully used to address the issue of bandwidth selection in the context of MCM (Chown et al., 2020; López-Cheda et al., 2017a,b). A review of bandwidth selection methods under the MCM can be found in López-Cheda (2018). In this thesis, the focus is on the bootstrap bandwidth selector.

1.5 Mixture cure model when the cure status is partially available

1.5.1 Model notation

Let Y be the time-to-event of interest. When there is a cured fraction, the survival function of Y is $S(t \mid x)$ in (1.4), so $Y = \infty$ with probability $1 - p(x)$, and the survival function of Y is $S_0(t \mid x)$ otherwise. Let $\nu = \mathbf{1}(Y = \infty)$ be the cure indicator. In standard cure models, the cure status is only known for a subject who experienced the event during the follow-up period ($\delta = 1$) and thus known to be uncured ($\nu = 0$). For a subject with $\delta = 0$ (censored) the cure status is unknown, thus $\nu = 1$ is never observed. To accommodate the possible availability of the cure status information, let the censoring distribution be an improper distribution function $G(t \mid x) = (1 - \pi(x)) G_0(t \mid x)$, so with probability $\pi(x)$ the censoring variable is $C = \infty$, and with probability $1 - \pi(x)$ the value of the censoring variable C corresponds to a random censoring time C_0 with proper continuous distribution function $G_0(t \mid x)$.

Hereto, a cured subject ($Y_i = \infty$), whose observed lifetime is always censored, is known to be cured if $C_i = \infty$. As a consequence, the cure status $\nu_i = 1$ is observed for some censored individuals. Clearly, a cured individual is identified with probability $\pi(x)$. Let ξ be a binary random variable which indicates whether the cure status ν is observed ($\xi = 1$) or not ($\xi = 0$). The conditional probability of observing the cure status is

$$\begin{aligned}P(\xi = 1 \mid X = x) =& P(\xi = 1 \mid Y < \infty, X = x)P(Y < \infty \mid X = x)\\ &+ P(\xi = 1 \mid Y = \infty, X = x)P(Y = \infty \mid X = x)\\ =& P(C < Y \mid Y < \infty, X = x)P(Y < \infty \mid X = x)\\ &+ P(C = \infty \mid Y = \infty, X = x)P(Y = \infty \mid X = x).\end{aligned}$$

If Y and C are independent conditionally on $X = x$, then the probability of observing the cure status is

$$P(\xi = 1 \mid X = x) = P(C < Y \mid X = x)p(x) + \pi(x)(1 - p(x)). \tag{1.7}$$

The observed data is $\{(X_i, T_i, \delta_i, \xi_i\nu_i), i = 1, \ldots, n\}$ and it is classified into three groups:

(a) when the individual is observed to have experienced the event, therefore, known to be uncured $(X_i, T_i, \delta_i = 1, \xi_i\nu_i = 0)$;

(b) when the lifetime is censored and the cure status is unknown $(X_i, T_i, \delta_i = 0, \xi_i \nu_i = 0)$;

(c) when the lifetime is censored and the individual is known to be cured $(X_i, T_i, \delta_i = 0, \xi_i \nu_i = 1)$,

where

$$T_i = \min(Y_i, C_i)\left[1 - \mathbf{1}(Y_i = \infty, C_i = \infty)\right] + C_{0i}\mathbf{1}(Y_i = \infty, C_i = \infty). \tag{1.8}$$

In the present context, when the observed times of the individuals known to be cured are random, the random variable C_0 models these observed *cured times.* In the unconditional setting in Betensky and Schoenfeld (2001), besides the lifetime Y and the censoring variable C, the so-called variable U plays a similar role. In standard cure models when the cure status is unknown for all the censored observations ($\pi(x) = 0$ and $C_i < \infty$, $\xi_i \nu_i = 0$, for $i = 1, \ldots, n$), then $T_i = \min(Y_i, C_i)$ and only groups (a) and (b) are considered.

To further understand the relationship between the notation introduced above and the usual notation in survival analysis under right censoring, let T_i in (1.8) be the *actual* observed times, and let $\widetilde{T}_i = \min(Y_i, C_i)$ denote the *usual* observed time, as it is usually defined in standard survival analysis. Note that $\widetilde{T}_i = T_i$ for the observations in groups (a) and (b). But if an observation is known to be cured ($\xi_i \nu_i = 1$), then $Y_i = \infty$ and $C_i = \infty$, and the *usual* observed time is $\widetilde{T}_i = \infty$. Nonetheless, the *actual* observed time for the individuals known to be cured is always finite and is recorded as $T_i = C_{0i}$. Therefore, when an individual is known to be cured it is guaranteed to observe a cure time C_{0i}, similarly to Betensky and Schoenfeld (2001).

In summary, if the observed times T_i of the censored individuals known to be cured are replaced by an extremely large observed time, say infinity, we recover the observations using the usual definition as $\tilde{T}_i = \min(Y_i, C_i)$.

1.6 Motivating examples

In the first example, a dataset of 233 sarcoma patients is studied. It includes patients who are $20 - 90$ years old from the University Hospital of Santiago de Compostela, Spain (CHUS). Sarcoma is a rare type of cancer that represents 1% of all adult solid malignancies. If a tumor can be surgically removed to render the patient with sarcoma free of detectable disease, 5 years is the survival time at which sarcoma oncologists assume long-term remissions (Choy, 2014). Overall, 59 patients died from sarcoma, and the remaining 174 patients were censored. Among censored patients, 18 patients were tumor free for more than five years. Hence, they were assumed to be long-term survivors. The aim was to estimate the survival time of the patients until death from sarcoma as a function of covariates such as the age at diagnosis, sex, tumor site, cancer spread (metastasis) and the margin status. The variables selected for estimating the survival probabilities were previously reported to be related to long-term sarcoma survival (Daigeler et al., 2014; Carbonnaux et al., 2019, among others).

The second example relates to a dataset of patients with breast cancer from The Cancer Genome Atlas (2021). A total of 898 breast cancer female patients were diagnosed and followed over time between 1988 and 2013. Information on demographic and clinical characteristics was collected at baseline. The goal of our analysis was to estimate the probability of not dying from breast cancer given specific characteristics of the patient. That is, the event of interest is death from breast cancer. The observed time-to-event was considered as censored if the event was not observed within the follow-up time, specifically, if the patient was alive at the end of the study, cancer-free either alive or dead, or lost to follow-up. Patients who have been free of cancer for at least 10 years can be considered to be free from dying of breast cancer (Pan et al., 2017; Barnadas et al., 2018). In our context, these patients are assumed to be cured, as they will not experience death from breast cancer. Therefore, they can be classified as long-term survivors or “known cured observations”. Note that the observed times-to-event of these “cured” individuals are very large, at least 10 years.

The third example considers the $n = 2,484$ COVID-19 patients hospitalized in Galicia (North-West of Spain) during the first few weeks of the outbreak. The data was collected by the Galician Healthcare Service (2021). The aim with this database is not so related to a medical goal like for the previous real data examples, rather with hospital management. During the first weeks of the pandemic, it was extremely necessary to plan and properly estimate the occupancy of both ward beds and ICU occupancy, in order to avoid overloads of the Galician healthcare system. To do this, it was basic to model the length of stay of these patients in hospital capacities. In particular, it was desired to know the probability that a patient admitted to the hospital would finally need to be admitted to the ICU, and the time it would take for these patients to be admitted to the ICU. That is, the event of interest was admission to ICU of hospital inpatients. The observed time-to-event was censored if the patient did not enter ICU during the follow-up time. Some of these censored patients were discharged or died before admission to ICU so they would never require ICU admission any more. In our context, this means that they would never experience the event of interest and can be considered as cured from the event. Unlike the previous examples, now the observed times-to-event of these “cured” individuals do not have to be necessarily large, as patients may die or be discharged after a short stay in the hospital.

Chapter 2

Generalized product-limit estimator of the survival function

2.1 Introduction

In this chapter, a generalized PL estimator of the survival function in the MCM when the cure status is partially known is proposed. It is based on a novel kernel estimator for the conditional cumulative hazard function. The rest of the chapter is organized as follows. In Section 2.2, (sub)distribution functions are defined and assumptions needed to construct theoretical results are stated. In Section 2.3, the estimators are defined and the asymptotic results are presented. Moreover, the bootstrap method when the cure status is partially known is described and a bootstrap procedure for bandwidth selection is proposed. Section 2.4 summarizes the results of a simulation study. An application to the sarcoma data and a discussion are provided in Sections 2.5 and 2.6, respectively.

2.2 Definitions and assumptions

Let us define the following (sub)distribution functions:

$$
\begin{aligned}
H\left(t \mid x\right) &= P\left(T \leq t \mid X = x\right), && (2.1)\\
H^{1}(t \mid x) &= P\left(T \leq t, \delta = 1 \mid X = x\right), && (2.2)\\
H^{11}(t \mid x) &= P\left(T \leq t, \xi\nu = 1 \mid X = x\right), && (2.3)\\
H^{0}(t \mid x) &= P\left(T \leq t, \xi\nu = 0 \mid X = x\right), && (2.4)\\
J(t \mid x) &= 1 - H\left(t \mid x\right) + H^{11}\left(t \mid x\right). && (2.5)
\end{aligned}
$$

The functions $H(t \mid x)$ and $H^1(t \mid x)$ are the conditional distribution function of the observed times T_i, and the conditional subdistribution function of the observed events, respectively. Note that $H^0(t \mid x)$ corresponds to the subdistribution function of the observed times of the individuals not known to be cured, and it reduces to the conditional distribution function $H(t \mid x)$ in the usual MCM when no individuals are known to be cured. Meanwhile, $H^{11}(t \mid x)$ provides insight about the distribution of *cure times*, that is, the recorded times T_i of the individuals known to be cured. Finally, $J(t \mid x)$ is the survival function of the observed times defined with the *usual* definition, that is, the conditional survival function of $\widetilde{T} = \min(Y, C)$.

The NW kernel estimators of (2.1) – (2.5) are, respectively,

$$\begin{aligned}
\widehat{H}_h\left(t \mid x\right) &= \sum_{i=1}^{n} B_{hi}(x)\mathbf{1}\left(T_i \leq t\right), && (2.6)\\
\widehat{H}_h^1(t \mid x) &= \sum_{i=1}^{n} B_{hi}(x)\mathbf{1}\left(T_i \leq t, \delta_i = 1\right), && (2.7)\\
\widehat{H}_h^{11}\left(t \mid x\right) &= \sum_{i=1}^{n} B_{hi}(x)\mathbf{1}\left(T_i \leq t, \xi_i\nu_i = 1\right), && (2.8)\\
\widehat{H}_h^{0}\left(t \mid x\right) &= \sum_{i=1}^{n} B_{hi}(x)\mathbf{1}\left(T_i \leq t, \xi_i\nu_i = 0\right), && (2.9)\\
\widehat{J}_h(t^- \mid x) &= \sum_{i=1}^{n} B_{hi}(x)\mathbf{1}\left(T_i \geq t\right) + \sum_{i=1}^{n} B_{hi}(x)\mathbf{1}\left(T_i < t, \xi_i\nu_i = 1\right). && (2.10)
\end{aligned}$$

In order to justify our asymptotic results, the following assumptions will be required (see similar assumptions in Iglesias-Pérez and González-Manteiga (1999)).

Assumption 1. (i) Let $I = [x_1, x_2]$ be an interval contained in the support of the density function of X, $m(x)$, such that

$$0 < \gamma = \inf_{x \in I_\varepsilon} m(x) < \sup_{x \in I_\varepsilon} m(x) = \Gamma < \infty$$

for some $I_\varepsilon = [x_1 - \varepsilon, x_2 + \varepsilon]$ with $\varepsilon > 0$ and $0 < \varepsilon\Gamma < 1$.

(ii) There exist $a, b \in \mathbb{R}$, $a < b$, satisfying $J\left(t \mid x\right) \geq \theta > 0$ for $(t, x) \in [a, b] \times I_\varepsilon$.

Assumption 2. (i) The first derivative with respect to x of $m(x)$ exists and is continuous in $x \in I_\varepsilon$.

(ii) The first derivatives with respect to x of $H(t \mid x)$, $H^1(t \mid x)$ and $H^{11}(t \mid x)$ exist and are continuous and bounded in $(t, x) \in [0, \infty) \times I_\varepsilon$.

Assumption 3. (i) The second derivative with respect to x of $m(x)$ exists and is continuous in $x \in I_\varepsilon$.

(ii) The second derivatives with respect to x of $H(t \mid x)$, $H^1(t \mid x)$ and $H^{11}(t \mid x)$ exist and are continuous and bounded in $(t, x) \in [0, \infty) \times I_\varepsilon$.

Assumption 4. The first derivatives with respect to t of $H(t \mid x)$, $H^1(t \mid x)$ and $H^{11}(t \mid x)$ exist and are continuous in $(t, x) \in [a, b] \times I_\varepsilon$.

Assumption 5. The second derivatives with respect to t of $H(t \mid x)$, $H^1(t \mid x)$ and $H^{11}(t \mid x)$ exist and are continuous in $(t, x) \in [a, b] \times I_\varepsilon$.

Assumption 6. The first derivative with respect to x and the second derivative with respect to t of $H(t \mid x)$, $H^1(t \mid x)$ and $H^{11}(t \mid x)$ exist and are continuous in $(t, x) \in [a, b] \times I_\varepsilon$.

Assumption 7. The (sub)densities corresponding to the (sub)distribution functions $H(t \mid x)$, $H^1(t \mid x)$ and $H^{11}(t \mid x)$ are bounded away from 0 in $[a, b] \times I_\varepsilon$.

Assumption 8. The kernel function $K(v)$ is a symmetrical density with zero mean, vanishing outside $(-1, 1)$, and the total variation is less than $\lambda < \infty$.

Assumption 9. The bandwidth $h = (h_n)$ satisfies $h \to 0, \log n/nh \to 0$, and $nh^5/\log n = O(1)$.

Assumption 10. (i) The fourth derivative with respect to x of $m(x)$ exists and is continuous in $x \in I_\varepsilon$.

(ii) The function $\pi(x)$ has at least two bounded derivatives.

Assumption 11. The bandwidths h_1, h_2 satisfy $((\log n)^3/nh_i)(h_j/(h_1+h_2))^2 \to 0$, for $i, j = 1, 2$, $i \neq j$, as $n \to \infty$.

2.3 Proposed estimator of the survival function

In this section, the proposed estimator of the conditional survival function in the MCM when the cure status is partially known is introduced. This estimator is based on the corresponding estimator of the cumulative hazard function $\Lambda(t \mid x)$. The main idea behind the construction of an estimator for $\Lambda(t \mid x)$ is the following. Note that the cumulative hazard function $\Lambda(t \mid x)$ can be written as follows:

$$\Lambda(t \mid x) = \int_0^t \frac{dF(v \mid x)}{1 - F(v^- \mid x)} = \int_0^t \frac{(1 - G(v^- \mid x))\, dF(v \mid x)}{(1 - G(v^- \mid x))(1 - F(v^- \mid x))}. \tag{2.11}$$

Observe that, if Y and C are conditionally independent given $X = x$, the denominator part of (2.11) is

$$\begin{aligned}
& \left(1 - F\left(v^- \mid x\right)\right)\left(1 - G\left(v^- \mid x\right)\right) \\
= \;& P(Y \geq v, C \geq v \mid X = x) \\
= \;& P(Y \geq v, C \geq v, \xi\nu = 0 \mid X = x) \\
& + P(Y \geq v, C \geq v, \xi\nu = 1 \mid X = x) \\
= \;& P(T \geq v, \xi\nu = 0 \mid X = x) + P(\xi\nu = 1 \mid X = x) \\
= \;& P(T \geq v \mid X = x) + P(T < v, \xi\nu = 1 \mid X = x) \\
= \;& 1 - H\left(v^- \mid x\right) + H^{11}\left(v^- \mid x\right).
\end{aligned} \tag{2.12}$$

On the other hand, for the numerator in (2.11) note that

$$\begin{aligned}
\int_0^t \left(1 - G\left(v^- \mid x\right)\right) dF(v \mid x) &= P(C \geq Y, Y \leq t \mid X = x) \\
&= P(T \leq t, \delta = 1 \mid X = x) \\
&= H^1(t \mid x).
\end{aligned} \tag{2.13}$$

By differentiating (2.13) and plugging it, together with (2.12), into (2.11), it is readily seen that

$$\Lambda(t \mid x) = \int_0^t \frac{dH^1(v \mid x)}{1 - H(v^- \mid x) + H^{11}(v^- \mid x)} = \int_0^t \frac{dH^1(v \mid x)}{J(v^- \mid x)}. \tag{2.14}$$

Now, by replacing in (2.14) $H^1(t \mid x)$ and $J(t \mid x)$ with their estimators (2.7) and (2.10), the following proposed estimator of $\Lambda(t \mid x)$ is obtained

$$\widehat{\Lambda}_h^c(t \mid x) = \sum_{i=1}^n \frac{\delta_{[i]} B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^n B_{h[j]}(x) + \sum_{j=1}^{i-1} B_{h[j]}(x) \mathbf{1}\left(\xi_{[j]}\nu_{[j]} = 1\right)}, \tag{2.15}$$

where $\delta_{[i]}$, $\xi_{[i]}$ and $\nu_{[i]}$ are the concomitants of the ordered observed times $T_{(1)} \leq \cdots \leq T_{(n)}$ and $B_{h[i]}(x)$ is defined in equation (1.3).

The distribution function corresponding to $\Lambda(t \mid x)$ is given by $S(t \mid x) = \exp(-\Lambda(t \mid x))$. After considering a Taylor's expansion of the exponential function around 0 and evaluating it at each increment of $\widehat{\Lambda}_h^c(t \mid x)$, the corresponding generalized product-limit estimator of the conditional survival function $S(t \mid x)$ when the cure status is partially known is

$$\widehat{S}_h^c(t \mid x) = \prod_{i=1}^n \left(1 - \frac{\delta_{[i]} B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^n B_{h[j]}(x) + \sum_{j=1}^{i-1} B_{h[j]}(x) \mathbf{1}\left(\xi_{[j]}\nu_{[j]} = 1\right)}\right). \tag{2.16}$$

In the next sections, the estimator in (2.16) is also referred to as $1 - \widehat{F}_h^c(t \mid x)$. It is readily seen that the proposed estimator for the censoring distribution $G(t \mid x)$ in the MCM when the cure status is partially known is

$$1 - \widehat{G}_h^c(t \mid x) = \prod_{i=1}^{n} \left(1 - \frac{\left(1 - \delta_{[i]}\right) \mathbf{1}\left(\xi_{[i]}\nu_{[i]} = 0\right) B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x) + \sum_{j=1}^{i-1} B_{h[j]}(x) \mathbf{1}\left(\xi_{[j]}\nu_{[j]} = 1\right)} \right). \quad (2.17)$$

In an unconditional setting, the estimator in (2.16) becomes (see the proof in Appendix A)

$$\widehat{S}_n^c(t) = \prod_{i=1}^{n} \left(1 - \frac{\delta_{[i]} \mathbf{1}\left(T_{(i)} \leq t\right)}{n - i + 1 + \sum_{j=1}^{i-1} \mathbf{1}\left(\xi_{[j]}\nu_{[j]} = 1\right)} \right). \quad (2.18)$$

An important feature of these estimators is that observations that are known to be cured before time $T_{(i)}$ remain in the risk set, i.e., they are counted in the denominator.

Proposition 2.1 The proposed estimator $\widehat{S}_h^c(t \mid x)$ has the following basic properties.

1. When there are no censored observations known to be cured, i.e., $\xi_i\nu_i = 0$ for $i = 1, \ldots, n$, $\widehat{S}_h^c(t \mid x)$ in (2.16) reduces to Beran's estimator in (1.2).

2. The survival function estimator in (2.16) is precisely Beran's estimator in (1.2) computed with the *usual* observed times $\left\{\left(\widetilde{T}_i, \delta_i\right), i = 1, \ldots, n\right\}$ where $\widetilde{T}_i = T_i$ if $\xi_i\nu_i = 0$ and $\widetilde{T}_i = \infty$ if $\xi_i\nu_i = 1$, that is, if the observed times of the individuals known to be cured are replaced with an extremely large value (e.g., infinity):
$$\widehat{S}_h^c(t \mid x) = \prod_{i=1}^{n} \left(1 - \frac{\delta_{[i]} B_{h[i]}(x) \mathbf{1}\left(\widetilde{T}_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x)} \right) \quad (2.19)$$
where $\delta_{[i]}$ and $X_{[i]}$ are the concomitants of the ordered *usual* observed times $\widetilde{T}_{(1)} \leq \cdots \leq \widetilde{T}_{(n)}$.

3. In the specific case when some individuals are considered as cured when their survival time exceeds a known fixed cure threshold, $\widehat{S}_h^c(t \mid x)$ also reduces to Beran's estimator in (1.2).

4. When there is no censoring, $\widehat{S}_h^c(t \mid x)$ reduces to the kernel estimator of the conditional survival function (Nadaraya, 1964b):
$$\widetilde{S}_h(t \mid x) = \sum_{i=1}^{n} B_{h[i]}(x) \mathbf{1}\left(Y_{(i)} > t\right).$$

5. In an unconditional setting and in the particular case where an individual is known to be cured only if the observed time is greater than a known fixed time, say d, $\widehat{S}_n^c(t)$ in (2.18) reduces to the generalized maximum likelihood estimator in Laska and Meisner (1992).

The proof of these properties is outlined in Appendix A.

Remark 2.1 The censoring distribution $G(t \mid x)$ can be estimated using the *usual* observations $\left\{\left(\widetilde{T}_i, \delta_i\right), i = 1, \ldots, n\right\}$ as follows:

$$1 - \widehat{G}_h^c(t \mid x) = \prod_{i=1}^{n} \left(1 - \frac{\left(1 - \delta_{[i]}\right) B_{h[i]}(x) \mathbf{1}\left(\widetilde{T}_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x)}\right) \tag{2.20}$$

where $\delta_{[i]}$ and $X_{[i]}$ are the concomitants of the ordered *usual* observed times $\widetilde{T}_{(1)} \leq \cdots \leq \widetilde{T}_{(n)}$.

Proposition 2.2 It can be shown that the distribution function $\widehat{H}_h^0(t \mid x)$ satisfies the relation

$$1 - \widehat{H}_h^0(t \mid x) = \left(1 - \widehat{F}_h^c(t \mid x)\right)\left(1 - \widehat{G}_h^c(t \mid x)\right)$$

where

$$1 - \widehat{H}_h^0(t \mid x) = \sum_{i=1}^{n} B_{hi}(x) \mathbf{1}(T_i > t, \xi_i \nu_i = 0)$$

is the kernel type estimator of $1 - H^0(t \mid x)$ in (2.4), and $1 - \widehat{F}_h^c(t \mid x)$ and $1 - \widehat{G}_h^c(t \mid x)$ are the estimators in (2.16) and (2.17), respectively.

The proof of Proposition 2.2 is given in Appendix A.

Proposition 2.3 The $1-\widehat{F}_h^c(t \mid x)$ estimator in (2.16) is the nonparametric local maximum likelihood estimator of $1 - F(t \mid x)$.

The proof of Proposition 2.3 is given in Appendix A.

2.3.1 Asymptotic results

In this section, the asymptotic properties of $\widehat{\Lambda}_h^c(t \mid x)$ and $\widehat{S}_h^c(t \mid x)$ are investigated. Theorems 2.1 and 2.2 below give the asymptotic representations of $\widehat{\Lambda}_h^c(t \mid x)$ and $1 - \widehat{F}_h^c(t \mid x)$, respectively. Corollary 2.1 shows that $\widehat{\Lambda}_h^c(t \mid x)$ and $1 - \widehat{F}_h^c(t \mid x)$ are strongly consistent estimators of $\Lambda(t \mid x)$ and $1 - F(t \mid x)$, respectively. The asymptotic normality of $1 - \widehat{F}_h^c(t \mid x)$ is proved in Theorem 2.3.

Theorem 2.1 (**Asymptotic representation of** $\widehat{\Lambda}_h^c(t \mid x)$) Suppose that Assumptions $1-9$ are satisfied. Then, for $x \in I$ and $t \in [a,b]$,

$$\widehat{\Lambda}_h^c(t \mid x) - \Lambda(t \mid x) = \sum_{i=1}^n \widetilde{B}_{hi}(x)\,\zeta(T_i, \delta_i, \xi_i, \nu_i, t, x) + R_{n1}(t,x),$$

where

$$\begin{aligned}&\zeta(T_i, \delta_i, \xi_i, \nu_i, t, x)\\ =&\frac{\mathbf{1}(T_i \le t, \delta_i = 1)}{J(T_i^- \mid x)} - \int_0^t \left(\mathbf{1}(T_i \ge v) + \mathbf{1}(T_i < v, \xi_i\nu_i = 1)\right)\frac{dH^1(v \mid x)}{J^2(v^- \mid x)},\end{aligned} \tag{2.21}$$

$$\widetilde{B}_{hi}(x) = \frac{1}{m(x)}\frac{1}{nh}K\left(\frac{x - X_i}{h}\right), \tag{2.22}$$

and $R_{n1}(t,x)$ satisfies

$$\sup_{a \le t \le b, x \in I} \mid R_{n1}(t,x) \mid = O\left((nh)^{-3/4}(\log n)^{3/4}\right) \quad \text{a.s.}$$

Theorem 2.2 (**Asymptotic representation of** $1 - \widehat{F}_h^c(t \mid x)$) Suppose that Assumptions $1-9$ hold. Then, for $x \in I$ and $t \in [a,b]$,

$$\widehat{F}_h^c(t \mid x) - F(t \mid x) = (1 - F(t \mid x))\sum_{i=1}^n \widetilde{B}_{hi}(x)\,\zeta(T_i, \delta_i, \xi_i, \nu_i, t, x) + R_{n2}(t,x),$$

where $\zeta(T_i, \delta_i, \xi_i, \nu_i, t, x)$ is defined in (2.21), $\widetilde{B}_{hi}(x)$ in (2.22) and $R_{n2}(t,x)$ satisfies

$$\sup_{a \le t \le b, x \in I} \mid R_{n2}(t,x) \mid \; = \; O\left((nh)^{-3/4}(\log n)^{3/4}\right) \text{ a.s.} \tag{2.23}$$

The sketch of the proofs of Theorems 2.1 and 2.2 is outlined in Appendix A. As an immediate consequence of these theorems, the following corollary on the strong consistency of the estimators $\widehat{\Lambda}_h^c(t \mid x)$ and $1 - \widehat{F}_h^c(t \mid x)$ is obtained.

Corollary 2.1 (**Strong consistency**) Suppose that Assumptions $1-9$ hold. Then, for $x \in I$ and $t \in [a,b]$,

$$\sup_{a \le t \le b, x \in I} \mid \widehat{\Lambda}_h^c(t \mid x) - \Lambda(t \mid x) \mid \; = \; O\left((nh)^{-1/2}(\log n)^{1/2}\right) \quad \text{a.s.}$$

and

$$\sup_{a \le t \le b, x \in I} \mid \widehat{F}_h^c(t \mid x) - F(t \mid x) \mid \; = \; O\left((nh)^{-1/2}(\log n)^{1/2}\right) \quad \text{a.s.}$$

The proof of Corollary 2.1 is outlined in Appendix A.

Proposition 2.4 (**Asymptotic bias and variance**) Suppose that Assumptions $1-9$ hold. Then, the bias and variance of the dominant term of $1-\widehat{F}_h^c(t \mid x)$ are, respectively,

$$\begin{aligned} \mu_{h,c}(t,x) &= h^2 B_c(t,x) + O\left(h^4\right), \\ \sigma^2_{h,c}(t,x) &= (nh)^{-1} s_c^2(t,x) + O(n^{-1}h), \end{aligned} \tag{2.24}$$

with

$$B_c(t,x) = \frac{(1-F(t \mid x))(2\Phi_c'(x,t,x)\, m'(x) + \Phi_c''(x,t,x)\, m(x)) d_K}{2m(x)}, \tag{2.25}$$

$$s_c^2(t,x) = \frac{(1-F(t \mid x))^2 \Phi_1^c(x,t,x)\, c_K}{m(x)}, \tag{2.26}$$

where $d_K = \int v^2 K(v) dv$ and $c_K = \int K^2(v) dv$. Besides, $\Phi_c'(y,t,x)$ and $\Phi_c''(y,t,x)$ are the first and second derivatives of $\Phi_c(y,t,x)$ with respect to y:

$$\begin{aligned} \Phi_c(y,t,x) &= \mathrm{E}\left(\zeta(T,\delta,\xi,\nu,t,x) \mid X=y\right) \\ &= \int_0^t \frac{dH^1(v \mid y)}{1-H(v^- \mid x) + H^{11}(v^- \mid x)} \\ &\quad - \int_0^t \frac{1-H(v^- \mid y) + H^{11}(v^- \mid y)}{\left(1-H(v^- \mid x) + H^{11}(v^- \mid x)\right)^2} dH^1(v \mid x). \end{aligned}$$

The expressions of $\Phi_c'(y,t,x)$ and $\Phi_c''(y,t,x)$ are given in Lemmas D.3 and D.4, respectively.

The expression of $\Phi_1^c(y,t,x)$ is given in Lemma D.5, which is

$$\Phi_1^c(x,t,x) = \mathrm{E}\left(\zeta^2(T,\delta,\xi,\nu,t,x) \mid X=x\right) = \int_0^t \frac{dH^1(v \mid x)}{\left(1-H(v^- \mid x) + H^{11}(v^- \mid x)\right)^2},$$

with $\zeta(T,\delta,\xi,\nu,t,x)$ given in (2.21).

The proof of Proposition 2.4 is outlined in Appendix A.

The following theorem, whose proof is in Appendix A, establishes the asymptotic normality of $1-\widehat{F}_h^c(t \mid x)$.

Theorem 2.3 (**Asymptotic normality**) Suppose that Assumptions $1-9$ are satisfied. For $x \in I$ and $t \in [a,b]$, it follows that:

(i) If $nh^5 \to 0$ and $(\log n)^3/nh \to 0$, then

$$(nh)^{1/2}\left(\widehat{F}_h^c(t \mid x) - F(t \mid x)\right) \xrightarrow{d} N(0, s_c^2(t,x)).$$

(ii) If $nh^5 \to C^5 > 0$, then

$$(nh)^{1/2}\left(\widehat{F}_h^c(t \mid x) - F(t \mid x)\right) \xrightarrow{d} N(C^{5/2} B_c(t,x), s_c^2(t,x)),$$

with $B_c(t,x)$ given in (2.25), $s_c^2(t,x)$ in (2.26) and C is a positive constant.

2.3.2 Effect of ignoring the cure status information

In this section a theoretical comparison between the estimator $1-\widehat{F}_h^c(t \mid x)$ in (2.16) and Beran's estimator in (1.2) is made. More precisely, in order to understand the effect of ignoring the cure status information, the dominant terms of the bias and variance of Beran's estimator are compared with those of the proposed estimator in Proposition 2.4.

The dominant terms of the asymptotic bias and variance of the Beran estimator are, respectively,

$$\begin{aligned} \mu_h(t,x) &= h^2 B(t,x) + O\left(h^4\right) && (2.27)\\ \sigma_h^2(t,x) &= (nh)^{-1} s^2(t,x) + O(n^{-1}h), && (2.28)\end{aligned}$$

with

$$\begin{aligned} B(t,x) &= \frac{(1-F(t\mid x))\left(2\Phi'(x,t,x)\,m'(x) + \Phi''(x,t,x)\,m(x)\right)d_K}{2m(x)}, && (2.29)\\ s^2(t,x) &= \frac{(1-F(t\mid x))^2\,\Phi_1(x,t,x)\,c_K}{m(x)}, && (2.30)\end{aligned}$$

where $\Phi'(y,t,x)$ and $\Phi''(y,t,x)$ are the first and the second derivatives of $\Phi(y,t,x)$ with respect to y (see Lemmas 4 and 5 in López-Cheda et al. (2017b)):

$$\begin{aligned} \Phi(y,t,x) &= \int_0^t \frac{dH^1(v\mid y)}{1-H(v^-\mid x)} - \int_0^t \frac{1-H(v^-\mid y)}{(1-H(v^-\mid x))^2}\,dH^1(v\mid x), && (2.31)\\ \Phi_1(x,t,x) &= \int_0^t \frac{dH^1(v\mid x)}{(1-H(v^-\mid x))^2}. \end{aligned}$$

The expressions (2.27) – (2.30) for Beran's estimator are equivalent to the bias and variance terms (2.24) – (2.26) for $\widehat{S}_h^c(t\mid x)$, replacing $\Phi_c(x,t,x)$ and $\Phi_1^c(x,t,x)$ with $\Phi(x,t,x)$ and $\Phi_1(x,t,x)$, respectively.

As for the variance, when the cure status information is ignored then $H^{11}(t\mid x)=0$ for all t and x. Therefore, $\Phi_1^c(x,t,x)\le \Phi_1(x,t,x)$.

Notice that, when the same bandwidth is used for both estimators, ignoring the cure status increases asymptotically the variance of the estimator.

Returning to the bias, by applying Lemma D.3, one has

$$\Phi_c'(x,t,x) = \Phi'(x,t,x) = -\frac{S'(t^-\mid x)}{S(t^-\mid x)},$$

where $S'(t\mid x)$ is the derivative of $S(t\mid x)$ with respect to x, meaning that the effect of knowing the cure status on the bias is given by $\Phi_c''(x,t,x)$. From Lemma D.4,

$$\Phi_c''(x,t,x) = 2\int_0^t \frac{G'(v^-\mid x)}{1-G(v^-\mid x)}\frac{d}{ds}\left(\frac{S'(s\mid x)}{S(s\mid x)}\right)\bigg|_{s=v^-} dv - \frac{S''(t^-\mid x)}{S(t^-\mid x)}, \quad (2.32)$$

where

$$G(t \mid x) = (1-\pi(x))G_0(t \mid x)$$

and $S'(t \mid x)$, $S''(t \mid x)$ and $G'(t \mid x)$ refer to the derivatives with respect to x. If the cure status is ignored, equation (2.32) reduces to

$$\Phi''(x,t,x) = 2\int_0^t \frac{G_0'(v^- \mid x)}{1-G_0(v^- \mid x)} \frac{d}{ds}\left(\frac{S'(s \mid x)}{S(s \mid x)}\right)\Bigg|_{s=v^-} dv - \frac{S''(t^- \mid x)}{S(t^- \mid x)}.$$

In terms of bias, the advantage of knowing the cure status is not straightforward as it depends on the derivatives of $\pi(x)$ and $G_0(t \mid x)$. This implies that there is no guarantee that there will be a gain in terms of bias for the proposed estimator with respect to Beran's estimator.

2.3.3 Bootstrap procedures when the cure status is partially known

The purpose of this section is to introduce two equivalent algorithms for bootstrapping with right censored data when the cure status is partially known. Our algorithms mimic the bootstrap ideas of Li and Datta (2001). First, the covariate X_i is resampled with replacement from $\{X_1,\ldots,X_n\}$ to obtain $\{X_1^*,\ldots,X_n^*\}$. Then, for each X_i^* the random variables $(T_i^*,\delta_i^*,\xi_i^*\nu_i^*)$ are generated following ideas similar to those in Li and Datta (2001), using the simple weighted bootstrap or the obvious bootstrap resampling methods.

The simple weighted bootstrap

Generate $\{X_1^*,\ldots,X_n^*\}$ from the empirical distribution of $\{X_1,\ldots,X_n\}$. Next, for each X_i^* generate $(T_i^*,\delta_i^*,\xi_i^*\nu_i^*)$ from the weighted empirical conditional distribution $\widehat{F}_g(t,d,z \mid X_i^*)$ given by:

$$\widehat{F}_g(t,d,z \mid X_i^*) = \sum_{j=1}^n B_{gj}(X_i^*)\mathbf{1}\left(T_j \le t, \delta_j \le d, \xi_j\nu_j \le z\right) \tag{2.33}$$

where $B_{gj}(x)$ are the NW weights (1.3) with bandwidth g.

The obvious bootstrap

Consider the estimators $1-\widehat{F}_g^c(t \mid x)$ in (2.16) and $\widehat{G}_g^c(t \mid x)$ in (2.17) of the survival function $1-F(t \mid x)$ and the censoring distribution $G(t \mid x)$ respectively, computed with the weights $B_{gi}(x)$ and the same bandwidth g. Simulate the bootstrap sample $\{(X_i^*,T_i^*,\delta_i^*,\xi_i^*\nu_i^*), i=1,\ldots,n\}$ as follows.

Step 1. Generate $\{X_1^*, \dots, X_n^*\}$ from the empirical distribution of $\{X_1, \dots, X_n\}$.

Step 2. For $i = 1, \dots, n$, set $Y_i^* = \infty$ with probability $1 - \widehat{F}_g^c \left(T_{(n)} \mid X_i^* \right) = 1 - \widehat{p}_g^c \left(X_i^* \right)$, and generate a finite survival time of a susceptible individual otherwise:

$$Y_i^* \sim \frac{1 - \widehat{F}_g^c \left(t \mid X_i^* \right) - \left(1 - \widehat{p}_g^c \left(X_i^* \right) \right)}{\widehat{p}_g^c \left(X_i^* \right)}.$$

Generate $C_i^* = \infty$ with probability $1 - \widehat{G}_g^c \left(T_{(n)} \mid X_i^* \right) = \widehat{\pi}_g^c \left(X_i^* \right)$, and

$$C_i^* \sim \frac{1 - \widehat{G}_g^c \left(t \mid X_i^* \right) - \pi_g^c \left(X_i^* \right)}{1 - \widehat{\pi}_g^c \left(X_i^* \right)} \qquad \text{otherwise.}$$

Let $\widehat{G}_{0g} \left(t \mid x \right)$ be the kernel estimator of $G_0 \left(t \mid x \right)$, the distribution function of the observed times of the individuals known to be cured:

$$\widehat{G}_{0g} \left(t \mid x \right) = \frac{\sum_{i=1}^n B_{gi} \left(x \right) \mathbf{1} \left(T_i \leq t, \xi_i \nu_i = 1 \right)}{\sum_{i=1}^n B_{gi} \left(x \right) \mathbf{1} \left(\xi_i \nu_i = 1 \right)}. \tag{2.34}$$

For each $i = 1, \dots, n$, generate C_{0i}^* from $\widehat{G}_{0g} \left(t \mid X_i^* \right)$. The bootstrap sample is $\left\{ \left(T_i^*, \delta_i^*, \xi_i^* \nu_i^* \right), i = 1, \dots, n \right\}$ where the bootstrap observed times are

$$T_i^* = \min \left(Y_i^*, C_i^* \right) \left[1 - \mathbf{1} \left(Y_i^* = \infty, C_i^* = \infty \right) \right] + C_{0i}^* \mathbf{1} \left(Y_i^* = \infty, C_i^* = \infty \right) \tag{2.35}$$

with

$$\delta_i^* = \mathbf{1} \left(Y_i^* < C_i^* \right), \tag{2.36}$$

$$\xi_i^* \nu_i^* = \mathbf{1} \left(Y_i^* = \infty, C_i^* = \infty \right). \tag{2.37}$$

Proposition 2.5 Assume there are no ties in the observed times $\{T_1, \dots, T_n\}$. Then, the simple weighted bootstrap and the obvious bootstrap are equivalent.

The proof of Proposition 2.5 is given in Appendix A.

The pilot bandwidth g should tend to 0 at a slower rate than the smoothing bandwidth h. This oversmoothing pilot bandwidth is required for the bootstrap integrated squared bias and variance to be asymptotically efficient estimators of the integrated squared bias and variance terms. With right censored data, Li and Datta (2001) recommend a local pilot bandwidth $g_x = c_x n^{-1/9}$ which coincides with the optimal order obtained by Cao and González-Manteiga (1993) for the uncensored case. Simulation results in Section 2.4 (see also López-Cheda et al. (2017a,b) for the usual MCM when cure status is not available) show that the choice of the pilot bandwidth has a small effect on the selected bootstrap bandwidth. We propose to use the same local pilot bandwidth as in López-Cheda et al. (2017a,b):

$$g_x = \frac{d_k^+(x) + d_k^-(x)}{2} 100^{1/9} n^{-1/9}, \tag{2.38}$$

where $d_k^+(x)$ and $d_k^-(x)$ are the distances from x to the kth nearest neighbor on the right and left, and k is a suitably chosen integer depending on the sample size. If there are not at least k neighbors on the right (or left), we use $d_k^+(x) = d_k^-(x)$ (or $d_k^-(x) = d_k^+(x)$). Following López-Cheda et al. (2017a,b), we suggest setting $k = [n/4]$.

2.3.4 Bootstrap bandwidth selection

In this section we introduce a bootstrap bandwidth selector to choose the smoothing parameter h of the proposed estimator $\widehat{S}_h^c(t \mid x)$. The bootstrap bandwidth, h_x^*, is the bandwidth minimizing the bootstrap version of the mean integrated squared error (MISE). This bootstrap MISE can be approximated using Monte Carlo by:

$$\mathrm{MISE}_x^*(h) \simeq \frac{1}{B} \sum_{b=1}^{B} \int \left(\widehat{S}_h^{c,*b}(v \mid x) - \widehat{S}_g^c(v \mid x) \right)^2 \omega(v,x) dv, \tag{2.39}$$

where $\widehat{S}_h^{c,*b}(t \mid x)$ is the proposed estimator computed with the bth bootstrap resample, $b = 1, \ldots, B$, and a bandwidth h, and $\widehat{S}_g^c(t \mid x)$ is the same estimator computed with the original sample and with a pilot bandwidth g. Note that $\omega(v,x)$ is a nonnegative weight function, intended to give lower weight in the right tail of the distribution. The algorithm to compute the bootstrap bandwidth for a fixed covariate value x, is as follows:

Step 1. With the original sample and the pilot bandwidth g, compute $\widehat{S}_g^c(t \mid x)$.

Step 2. Choose a dense enough grid of L bandwidths $\{h_1, \ldots, h_L\}$.

Step 3. Generate B bootstrap resamples $\{(X_i^{*(b)}, T_i^{*(b)}, \delta_i^{*(b)}, \xi_i^{*(b)} \nu_i^{*(b)}), i = 1, \ldots, n\}$, for $b = 1, \ldots, B$.

Step 4. For the bth bootstrap resample and the bandwidths h_l, for $l = 1, \ldots, L$, compute $\widehat{S}_{h_l}^{c,*b}(t \mid x)$.

Step 5. For $h_l, l = 1, \ldots, L$, compute the Monte Carlo approximation of $\mathrm{MISE}_x^*(h_l)$ given by (2.39).

Step 6. The bootstrap bandwidth, h_x^*, is the bandwidth of the grid $\{h_1, \dots, h_L\}$ that minimizes the approximation of $\mathrm{MISE}_x^*(h)$ in (2.39).

The bootstrap resamples in Step 3 are generated following any of the two equivalent resampling algorithms introduced in Section 2.3.3. For computational efficiency (see López-Cheda et al., 2017a,b), we fixed $X_i^* = X_i$ instead of resampling it randomly from $\{X_1, \dots, X_n\}$.

2.4 Simulation study

The practical performance of $\widehat{S}_h^c(t \mid x)$ was studied through a simulation study. The estimators considered for comparison are the Beran estimator, $\widehat{S}_h(t \mid x)$, which ignores the information of the cure status, and the semiparametric estimator, $S(t \mid x; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$, by Bernhardt (2016), which takes advantage of the cure status information and fits a logistic regression for the cure probability and an AFT model for the latency function.

Observations were simulated from the conditional survival function $S(t \mid x) = 1 - p(x) + p(x)S_{0,1}(t \mid x)$, where

$$S_{0,1}(t \mid x) = \begin{cases} \dfrac{\exp(-\alpha(x)t) - \exp(-\alpha(x)4.605)}{1 - \exp(-\alpha(x)4.605)} & 0 \le t \le 4.605 \\ 0 & t > 4.605 \end{cases}, \qquad (2.40)$$

with $\alpha(x) = \exp((x+20)/40)$.

Two scenarios given by the cure probabilities were considered:

$$1 - p_1(x) = 1 - \frac{\exp(0.476 + 0.358x)}{1 + \exp(0.476 + 0.358x)}, \qquad 1 - p_2(x) = 0.5 - \frac{1}{16000}x^3.$$

The covariate X was uniformly distributed on the interval $[-20, 20]$. The censoring variable C was generated from an exponential distribution with mean $10/3$ with probability $1 - \pi(x)$ and $C = \infty$ with probability $\pi(x)$. In both scenarios, the proportion of cured individuals that are identified was set to $\pi(x) = 0.2$ and $\pi(x) = 0.8$. The percentage of censoring was about 47% when $\pi(x) = 0.8$ and 48% when $\pi(x) = 0.2$ in Scenario 1. In Scenario 2, 51% and 52% of the observations were censored when $\pi(x) = 0.8$ and $\pi(x) = 0.2$ respectively. The average cure probability was 0.467 in Scenario 1 and 0.5 in Scenario 2. Data were generated such that the censoring times C and the lifetimes Y were independent conditionally on $X = x$. For both scenarios, 1000 datasets of sample sizes $n = 50, 100$ and 200 were generated. The Epanechnikov kernel was chosen to compute $\widehat{S}_h^c(t \mid x)$ and $\widehat{S}_h(t \mid x)$.

The first goal was to evaluate the performance of $\widehat{S}_h^c(t \mid x)$ in terms of MISE. MISE was approximated over a grid of bandwidths equispaced in a logarithmic scale, from $h_1 = 3$ to $h_{100} = 20$ in Scenario 1, and from $h_1 = 4$ to $h_{101} = 100$ in Scenario 2. For the weight function, $\omega(t, x) = \mathbf{1}(a_x \leq t \leq b_x)$ was chosen where $a_x = 0$ and $b_x = \tau_x$, the 90th percentile of $S_0(t \mid x)$. Note that the semiparametric estimator $S_0(t \mid x; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$ is expected to perform well in Scenario 1.

Figure 2.1 shows the MISE curves of the three estimators. In Scenario 1, as expected, $S(t \mid x; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$ behaves well. Nevertheless, both $\widehat{S}_h^c(t \mid x)$ and $\widehat{S}_h(t \mid x)$ are quite competitive for suitable values of the bandwidth, even beating the semiparametric estimator for some values of X close to 0 and 20. In Scenario 2, both $\widehat{S}_h^c(t \mid x)$ and $\widehat{S}_h(t \mid x)$ outperform $S(t \mid x; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$, as the parametric models assumed by the semiparametric estimator are not met. Regarding the nonparametric estimators, taking into account the known cure status gives either similar or better results than ignoring it for most values of X, especially in Scenario 2. Figure 2.1 also displays the effect of the sample size on the behavior of the estimators when $\pi(x) = 0.8$. As the sample size increases, the MISE of the three estimators decreases as expected, while the differences in performance seem to fade out.

In Table 2.1, the performance of the estimators is compared in terms of the integrated squared bias, integrated variance and MISE for the covariate values $x = -10, 0$ and 10. In both scenarios, at $x = -10$, the proposed estimator has smaller integrated squared bias and variance than Beran's estimator. On the contrary, for $x = 10$, the integrated squared bias and variance of Beran's estimator is smaller compared to $\widehat{S}_h^c(t \mid x)$ estimator. As expected, the integrated squared bias and variance estimates for the semiparametric estimator are larger in Scenario 2.

Figure 2.2 and Table 2.2 provide some insight about the effect of $\pi(x)$ on the estimators. It should be noted that when $\pi(x)$ increases the censoring percentage decreases. While the effect of $\pi(x)$ on $\widehat{S}_h(t \mid x)$ is not straightforward, the behavior of $\widehat{S}_h^c(t \mid x)$ and $S(t \mid x; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$ improves in general as $\pi(x)$ increases. The proposed estimator attains a notable improvement in performance compared to Beran's estimator when $\pi(x) = 0.8$ in Figure 2.1 as compared to $\pi(x) = 0.2$ in Figure 2.2. Similar results are observed in Table 2.2, where the performance of the estimators for the covariate values $x = -10, 0, 10$ is compared in terms of the integrated squared bias, integrated variance and MISE.

The performance of the bootstrap bandwidth selector was assessed using $B = 1000$ resamples and an increased grid of bandwidths from 1.5 to 100 for both scenarios. Figure 2.3 displays the quartile of the selected bootstrap bandwidths together with the optimal bandwidth. Corresponding contour plots in Figure 2.4 show the density of the bootstrap bandwidths and the MISE of $\widehat{S}_h^c(t \mid x)$ as a function of the bandwidth h and the covariate value x. Figure 2.5 shows the MISE of $\widehat{S}_h^c(t \mid x)$ as a function of the bandwidth h, for four values of the covariate. Figure 2.3 and Figure 2.4 illustrate that the bootstrap bandwidth approximates quite well the optimal bandwidth. Note that in Figure 2.4 vertical contour lines indicate that, given x, the MISE of $\widehat{S}_h^c(t \mid x)$ tends to be constant as a function of h. Therefore, different bandwidths would yield approximately the same MISE. In those cases, the bootstrap bandwidth being far from the optimal bandwidth does not imply a loss of efficiency. Similar results are observed in Figure 2.5. For example, when $x = 0$ in Scenario 2, it is evident that the MISE initially decreases as the bandwidth increases, although afterwards it becomes constant.

To illustrate the effect of the pilot bandwidth g on the selected bootstrap bandwidth h_x^*, some simulations with different values of the number of neighbors, k were performed. Figure 2.6 shows the results for $k = [n/3]$, $k = [n/4]$ and $k = [n/8]$ when $n = 50$, $\pi = 0.8$ and $B = 500$. In both scenarios, it is observed that changing the value of k when computing the pilot bandwidth does not have a strong effect on h_x^*.

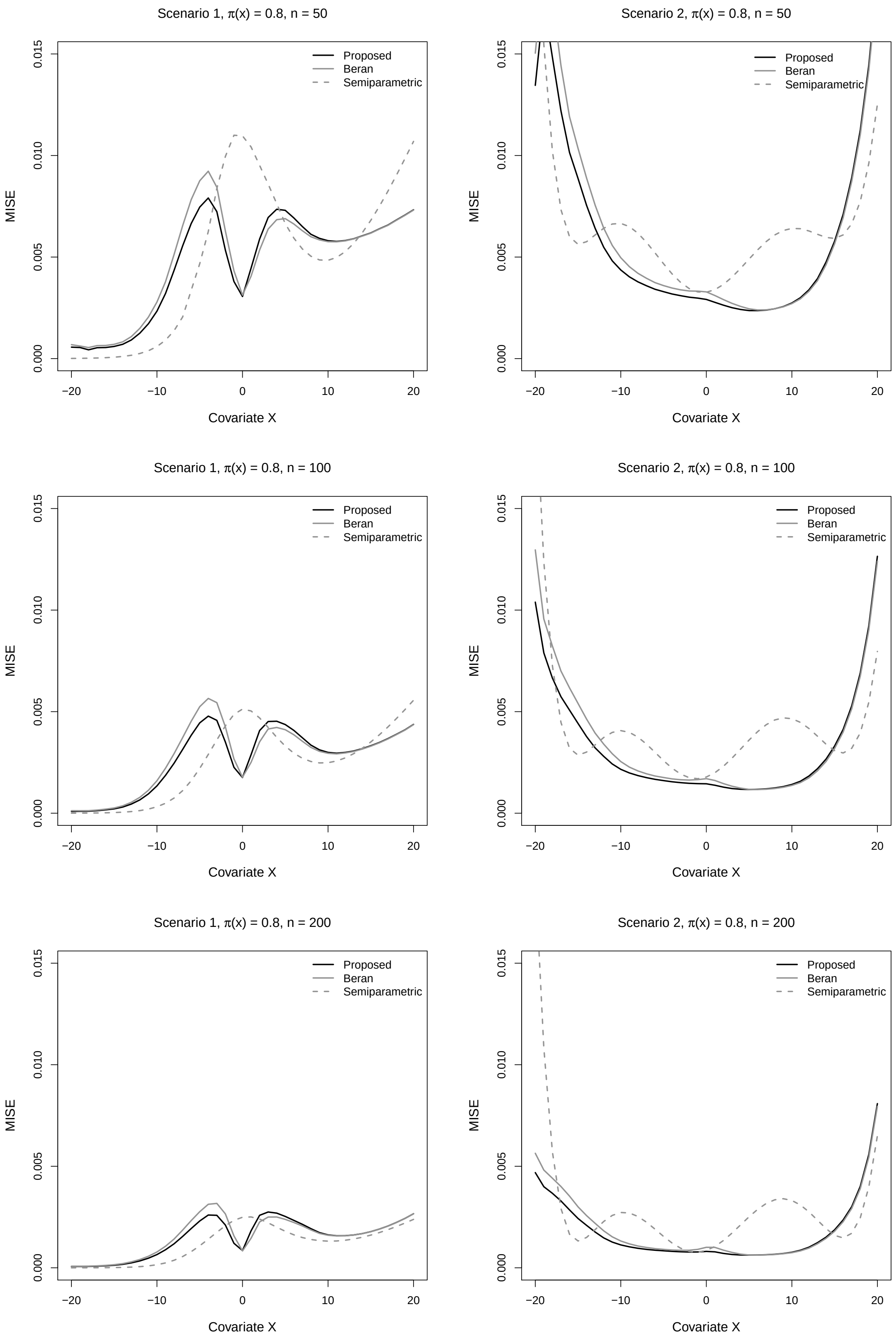

Figure 2.1: MISE of $\widehat{S}_h^c(t \mid x)$, $\widehat{S}_h(t \mid x)$ (both computed with the optimal bandwidth), and $S_0(t \mid x; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$ in Scenarios 1 (left) and 2 (right) for $\pi(x) = 0.8$, and $n = 50$ (top), 100 (center) and 200 (bottom).

Table 2.1: Integrated squared bias (Ibias^2), integrated variance (Ivar) and MISE of $\widehat{S}_h^c(t \mid x)$, $\widehat{S}_h(t \mid x)$ (both computed with the optimal bandwidth), and $S(t \mid x; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$ for $\pi(x) = 0.8$ and $n = 100$.

		$\widehat{S}_h^c(t \mid x)$				$\widehat{S}_h(t \mid x)$				$S(t \mid x; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$		
Scenario	x	h	Ibias^2 $\times 10^3$	Ivar $\times 10^3$	MISE $\times 10^3$	h	Ibias^2 $\times 10^3$	Ivar $\times 10^3$	MISE $\times 10^3$	Ibias^2 $\times 10^3$	Ivar $\times 10^3$	MISE $\times 10^3$
1	-10	5.998	0.135	1.211	1.346	5.998	0.201	1.399	1.600	0.000	0.316	0.317
	0	71.464	0.081	1.675	1.756	46.957	0.034	1.736	1.770	0.183	4.942	5.126
	10	11.744	0.242	2.752	2.994	12.247	0.261	2.688	2.949	0.221	2.265	2.486
2	-10	23.981	0.053	2.106	2.160	22.049	0.134	2.407	2.542	1.694	2.372	4.066
	0	28.368	0.092	1.357	1.449	25.010	0.258	1.442	1.700	0.137	1.645	1.782
	10	27.201	0.054	1.361	1.415	29.585	0.074	1.296	1.369	2.598	2.055	4.654

Table 2.2: Integrated squared bias (Ibias^2), integrated variance (Ivar) and MISE of $\widehat{S}_h^c(t \mid x)$, $\widehat{S}_h(t \mid x)$ (both computed with the optimal bandwidth), and $S(t \mid x; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$ in Scenario 1 for $\pi(x) = 0.2, 0.8$, and $n = 100$.

		$\widehat{S}_h^c(t \mid x)$				$\widehat{S}_h(t \mid x)$				$S(t \mid x; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$		
$\pi(x)$	x	h	Ibias^2 $\times 10^3$	Ivar $\times 10^3$	MISE $\times 10^3$	h	Ibias^2 $\times 10^3$	Ivar $\times 10^3$	MISE $\times 10^3$	Ibias^2 $\times 10^3$	Ivar $\times 10^3$	MISE $\times 10^3$
0.2	-10	5.998	0.140	1.290	1.431	5.998	0.157	1.339	1.495	0.001	0.332	0.333
	0	60.413	0.081	1.750	1.831	53.262	0.054	1.764	1.817	0.229	5.485	5.714
	10	11.744	0.248	2.849	3.097	11.744	0.238	2.852	3.090	0.088	2.278	2.365
0.8	-10	5.998	0.135	1.211	1.346	5.998	0.201	1.399	1.600	0.000	0.316	0.317
	0	71.464	0.081	1.675	1.756	46.957	0.034	1.736	1.770	0.183	4.942	5.126
	10	11.744	0.242	2.752	2.994	12.247	0.261	2.688	2.949	0.221	2.265	2.486

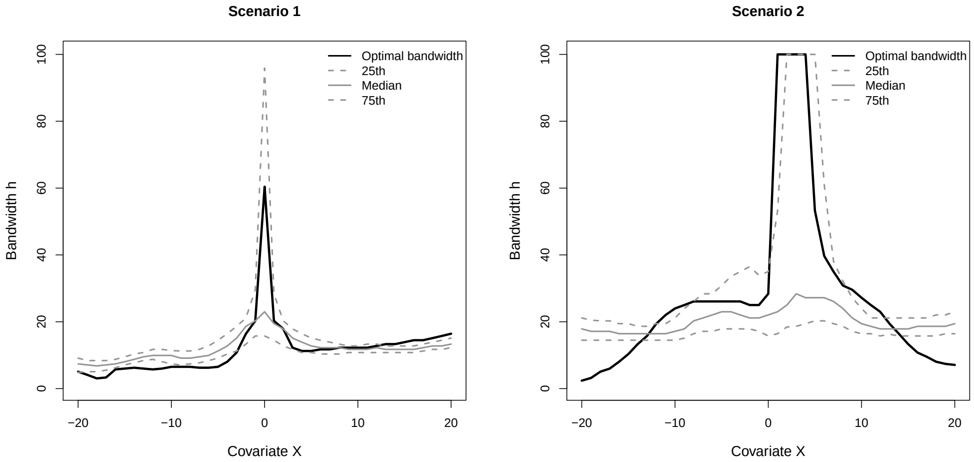

Figure 2.3: Median, first and third quartile of the bootstrap bandwidths for $\widehat{S}_h^c(t \mid x)$ in Scenarios 1 (left) and 2 (right) for $\pi(x) = 0.8$ and $n = 100$. The optimal bandwidth is displayed for reference.

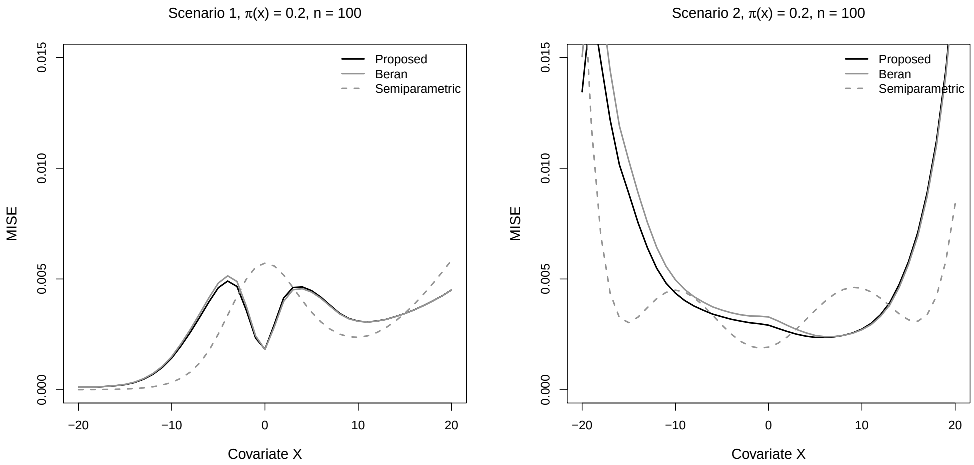

Figure 2.2: MISE of $\widehat{S}_h^c(t \mid x)$, $\widehat{S}_h(t \mid x)$ (both computed with the optimal bandwidth), and the semiparametric estimator in Scenarios 1 (left) and 2 (right) for $\pi(x) = 0.2$ and $n = 100$.

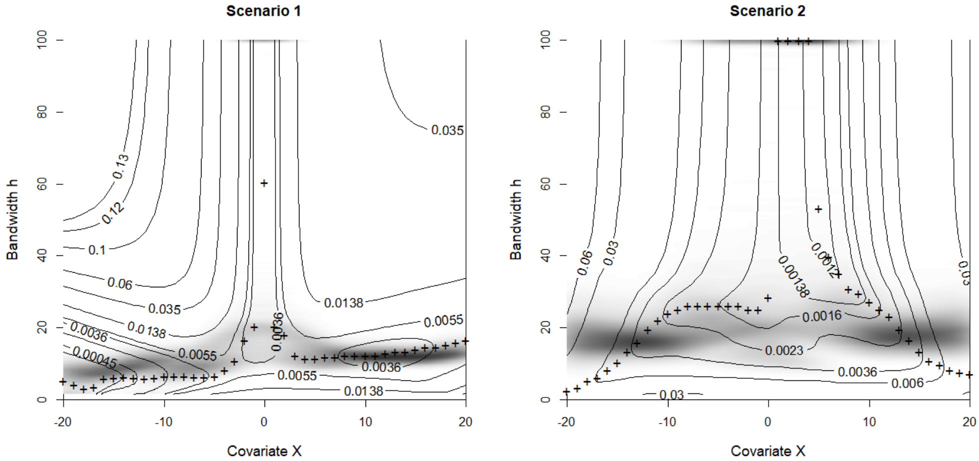

Figure 2.4: Contour plots of the MISE of $\widehat{S}_h^c(t \mid x)$ as a function of the bandwidth h and the covariate value x in Scenarios 1 (left) and 2 (right) for $\pi(x) = 0.8$ and $n = 100$. For each covariate value, the optimal bandwidth is marked with a cross. The density of the bootstrap bandwidths h_x^* is shown in gray shades (where a darker gray represents a higher density).

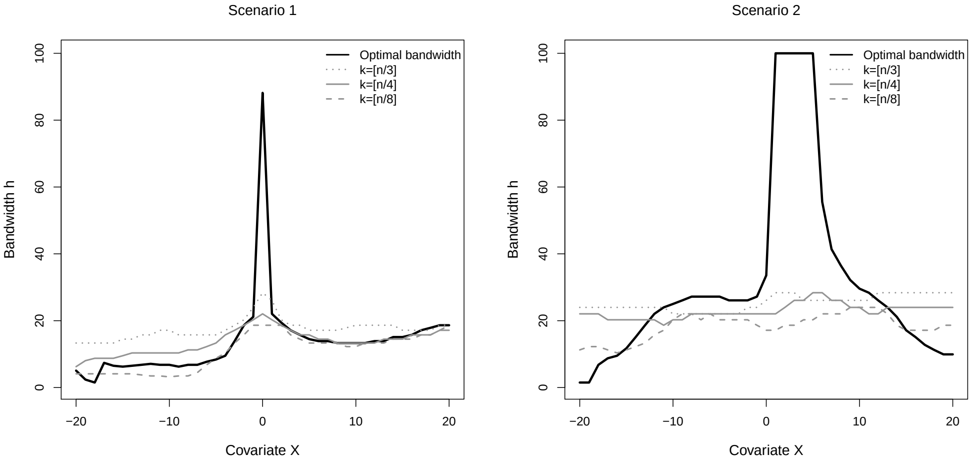

Figure 2.6: Median of the bootstrap bandwidths for $\widehat{S}_h^c(t \mid x)$ when using three different pilot bandwidths computed with $k = [n/3]$, $[n/4]$ and $[n/8]$ in Scenarios 1 (left) and 2 (right) for $\pi(x) = 0.8$ and $n = 50$. The optimal bandwidth is displayed for reference .

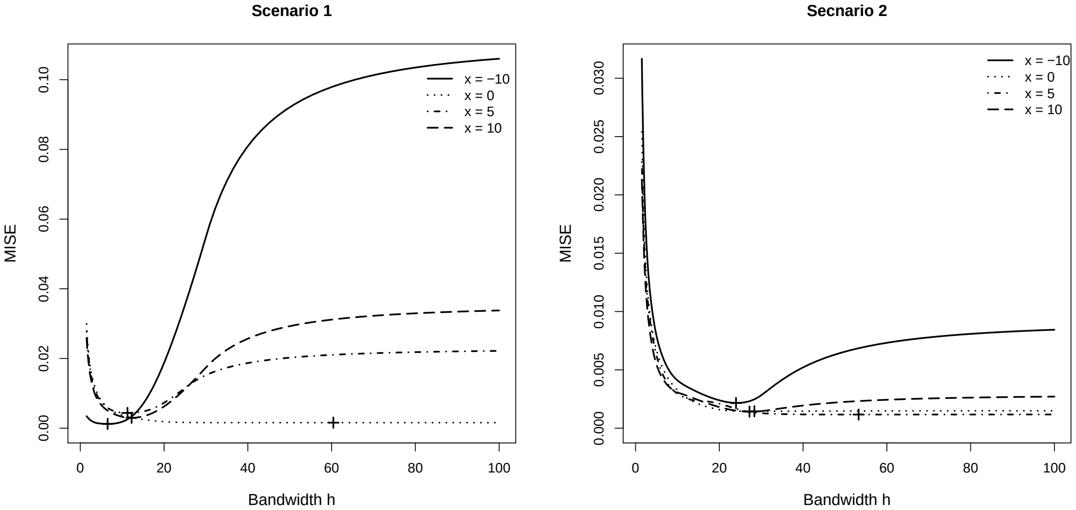

Figure 2.5: MISE of $\widehat{S}_h^c(t \mid x)$ as a function of the bandwidth h for covariate values $x = -10, 0, 5, 10$ in Scenarios 1 (left) and 2 (right) for $\pi(x) = 0.8$ and $n = 100$. For each value of the covariate, the optimal bandwidth is marked with a cross.

2.5 Real data analysis

The practical performance of the proposed estimators of the conditional survival function, $\widehat{S}_h^c(t \mid x)$ in (2.16) and its unconditional counterpart $\widehat{S}_n^c(t)$ in (2.18), is illustrated by applying the estimators to the sarcoma data. Table 2.3 shows the main demographic and clinical characteristics of the sarcoma patients.

Based on the asymptotic normality of the estimators, 95% confidence intervals (CI) of the conditional survival function have been constructed. Estimates of the standard error of the estimators have been obtained using the bootstrap procedure. More in detail, the $100(1-\alpha)\%$ CI of the (conditional) survival function with $\widehat{S}_h^c(t \mid x)$ and $\widehat{S}_n^c(t)$ were defined as, respectively,

$$\widehat{S}_h^c(t \mid x) \quad \mp \quad z_{1-\frac{\alpha}{2}} \widehat{se}_B \left(\widehat{S}_h^c(t \mid x) \right)$$

and

$$\widehat{S}_n^c(t) \quad \mp \quad z_{1-\frac{\alpha}{2}} \widehat{se}_B \left(\widehat{S}_n^c(t) \right),$$

z_β is the βth quantile of the standard normal, and $\widehat{se}_B \left(\widehat{S}_h^c(t \mid x) \right)$ and $\widehat{se}_B \left(\widehat{S}_n^c(t) \right)$ are, respectively, the bootstrap estimates of the standard errors of $\widehat{S}_h^c(t \mid x)$ and $\widehat{S}_n^c(t)$ computed with $B = 1000$ bootstrap resamples.

Figure 2.7 compares the estimates obtained with $\widehat{S}_h^c(t \mid x)$, which takes into account the 18 long-term survivors, those obtained with $\widehat{S}_h(t \mid x)$, which ignores individuals known to be cured and treats them as simply censored observations. Both estimators were computed using the corresponding bootstrap bandwidth selector introduced in Section 2.3.3 using $B = 1000$ resamples. The semiparametric estimator $S(t \mid x; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$ was also considered for reference. All estimators show that the survival curve decreases when age increases from 40 to 90 years. The largest differences between $\widehat{S}_h^c(t \mid x)$ and $\widehat{S}_h(t \mid x)$ were found at the right tail of the distribution, where the survival curve for $\widehat{S}_h^c(t \mid x)$ is slightly higher. Since the cure probability can be obtained as the limit of $S(t \mid x)$ when $t \to \infty$ using the proposed estimator of the survival curve will yield in higher estimates of the probability of cure.

On the other hand, the survival curve estimated by $S(t \mid x; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$ tends to decrease much slower than those obtained with $\widehat{S}_h^c(t \mid x)$ and $\widehat{S}_h(t \mid x)$, suggesting that further testing is required to provide evidence that assumptions in the semiparametric model are fulfilled.

Figure 2.7 at the bottom shows the survival curves of the sarcoma patients stratified by margin status. In this case, the proposed estimator in an unconditional

Table 2.3: Demographic and clinical characteristics of sarcoma patients. Also given are the numbers of died patients (Death), patients known to be cured (Cured) and patients with unknown cure status (Unknown).

Characteristic	n (%)	**Death**	**Censored**	
			Cured	**Unknown**
Age†				
< 60	105 (45.3%)	25	9	71
≥ 60	127 (54.7%)	33	9	85
Sex				
Male	100 (42.9%)	25	7	68
Female	133 (57.1%)	34	11	88
Tumor site†				
Retroperitoneal	86 (37.2%)	28	4	54
Extremities	70 (30.3%)	14	5	51
Other sites	75 (32.5%)	16	9	50
Metastatic†				
No	112 (67.1%)	11	9	92
Yes	55 (32.9%)	32	3	20
Margin status†				
Negative	133 (65.8%)	26	12	95
Positive	69 (34.2%)	17	3	49

† Contains a few missing data.

setting, $\widehat{S}_n^c(t)$, is applied and the KM estimator is considered as reference. The survival curves tend to decrease with time in both subgroups. The positive margin survival curve decreases slightly faster than the negative margin survival curve. In addition, the distinction between $\widehat{S}_n^c(t)$ and the KM estimator is found at the right tail of the distribution with the survival curves estimated with $\widehat{S}_n^c(t)$ being slightly higher than the KM curves. For example, the survival probability, at the tail of the distribution, for patients with negative margins is around 0.51 when estimated by $\widehat{S}_n^c(t)$, while it is around 0.47 when estimated with the KM estimator. Again, the estimated probability of cure is slightly higher when the survival curve is fitted taking into account the known cured subjects.

2.6 Discussion

The proposed estimator of the survival function takes advantage of the additional cure status information that Beran's estimator ignores, by considering the indi-

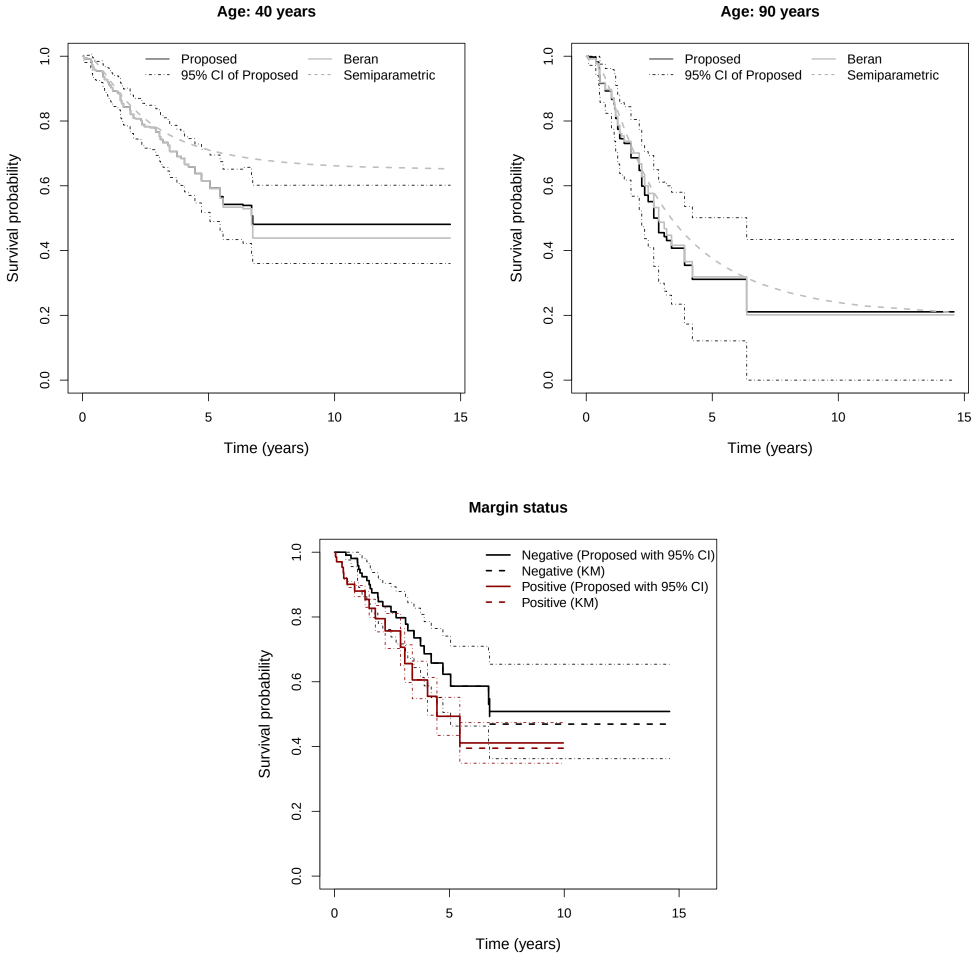

Figure 2.7: (Top panels) Survival estimates for sarcoma patients aged 40 and 90 years obtained using $\widehat{S}^c_h(t \mid x)$ and its 95% CI, $\widehat{S}_h(t \mid x)$, both computed with the bootstrap bandwidth, and $S(t \mid x; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$. (Bottom panel) Survival estimates stratified obtained using $\widehat{S}^c_n(t)$ with its 95% CI and $\widehat{S}_n(t)$.

viduals known to be cured always in the risk set, regardless of the values of their observed times. This is tantamount to shifting the observed times T_i of the individuals known to be cured arbitrarily to the right (in the limit, $\widetilde{T}_i = \infty$) and computing Beran's estimator with these modified observed times. If the observed times T_i of all these cured individuals are very large, e.g., when a subject is considered cured when the observed time surpasses a cure threshold, shifting these times T_i arbitrarily to the right has no effect, and the estimation of the survival function with the proposed estimator or with Beran's estimator ignoring the cure status is equivalent.

Although the proposed estimator utilizes the cure status information and shows good results both theoretically and practically, it is not without limitations. It is competitive over Beran's estimator in terms of the MISE, showing a generally better behavior. But when the sample size is small, the simulation results showed that for some values of the covariate it does not result in an improvement, but, a slightly worse performance in terms of MISE. The clear gain in terms of the integrated variance could be cancelled out by the integrated squared bias, which depends on the conditional probability of individuals identified as being cured and the conditional censoring distribution.

The performance of the semiparametric estimator by Bernhardt (2016) strongly depends on the plausibility of the parametric models assumed for either the cure rate or the latency fuction. Besides, our numerical experience indicates that if the sample size is small (less than 100), it is challenging to obtain stable estimates for the model parameters.

Approaches to include the cure information in the estimation of the MCM are scarcely developed in the literature. Many of them are based on a cure threshold, in which an individual is assumed to be cured from the event when the observed time-to-event surpasses a threshold, that is, the event was not observed for a fixed period time (see, e.g., Laska and Meisner, 1992; Nieto-Baraja and Yin, 2008). The proposed methodology not only encompasses this setup, but also generalizes it, as it can be applied when a subject is considered cured based on external information, that is, when the "cured" censored times are random.

Chapter 3

Kernel estimator of the cure probability

3.1 Introduction

When there is evidence of existence of long-term survivors and the cure status is known for some cured individuals, it is often of interest to estimate the probability of cure, which is the complementary of the probability of the event of interest. To properly estimate the cure probability when the cure status is partially available, a kernel estimator of the conditional cure probability based on the MCM is proposed in this chapter. This estimator extends the XP estimator to incorporate the cure status information.

A different approach for modeling survival data with a cure fraction when cure is randomly observed is a competing risks model, in which the event of interest and being identified as cured are competing risks failures, and only the minimum of the times of the corresponding risks is observed (Betensky and Schoenfeld, 2001; Nicolaie et al., 2019). The probability of cure is simply the cumulative incidence function of the competing risk given by cure, evaluated at the last observed time. The previous approaches for estimating the cure rate are methodologies where the time-to-event is of interest. Alternatively, note that the cure probability can be written as a function of a covariate X and it can be regarded as the conditional expectation of the cure status. As a consequence, standard regression techniques can be used to model the probability of cure, whereby these methods are relatively simple to implement. Since the survival times are subject to right censoring, the cure status is partially observed. In fact, denoting ν the cure status, $\nu = 0$ when the event is observed, $\nu = 1$ when cure is observed, and ν is missing for some censored observations. In the latter case, the individual will either experience the event (the observation is censored and susceptible) or never experience the event (the observation is censored and non-susceptible in the future). Therefore, this turns the estimation of the cure probability into a regression problem with missing response values. The most commonly used regression-based techniques to deal with missing data are inverse probability weighting (IPW) (Horvitz and Thompson, 1952; Robins et al., 1994; Lipsitz et al., 1998, among many others) and multiple imputation (MI) (Aerts et al., 2002; Rubin, 2004; Carpenter and Kenward, 2012; Wei et al., 2012). In the IPW approach (Seaman and White, 2013), only complete observations are included in the analysis, and weights are used to adjust the set of complete observations so it is representative of the entire sample. In the MI approach (Seaman et al., 2012), missing observations are replaced by values that are randomly drawn from the observed data, given some sampling scheme. Then, one can perform standard regression with the imputed data. Commonly used imputation methods for missing response values include semiparametric imputation (Wang et al., 2004), nearest neighbor imputation (Andridge and Little, 2010), and kernel-based techniques (Aerts et al., 2002; Cheng, 1994; Hsu et al., 2016). Aerts et al. (2002) applied the nonparametric kernel regression imputation scheme to estimate the unconditional mean. In this chapter, an estimator of the conditional cure probability based on a regression fit with multiple imputation for the unknown cure status is introduced as an alternative to the proposed MCM based nonparametric estimator. Note that the aforementioned imputation methods are strongly dependent on the proportion of missing data, giving unreliable estimations when there is a substantial level of missingness.

The rest of the chapter is organized as follows. In Section 3.2, a kernel estimator of the cure probability based on the MCM is proposed and its asymptotic properties are studied. Also, a bootstrap bandwidth selector is proposed. In Section 3.3, alternative estimators of the cure rate based, respectively, on the competing risks and nonparametric multiple imputation approaches are presented. The performance of these estimators is illustrated with a simulation study in Section 3.4. The estimators are applied to the breast cancer and COVID-19 datasets in Section 3.5. Finally, a discussion is provided in Section 3.6.

3.2 Proposed estimator of the cure rate

The cure rate is the probability that the event will not happen:

$$1 - p(x) = P(Y = \infty \mid X = x) = \lim_{t \to \infty} P(Y > t \mid X = x) = \lim_{t \to \infty} S(t \mid x).$$

A suitable estimator of the cure rate $1 - p(x)$ could easily be derived as the limit as t tends to infinity of an estimator of $S(t \mid x)$. Starting from the generalized product-limit estimator in (2.16), the following estimator for the cure probability, $1 - p\,(x)$, when the cure status is partially known is proposed:

$$1 - \widehat{p}_h^c\,(x) = \prod_{i=1}^{n} \left(1 - \frac{\delta_{[i]} B_{h[i]}\,(x)}{\sum_{j=i}^{n} B_{h[j]}\,(x) + \sum_{j=1}^{i-1} B_{h[j]}\,(x)\,\mathbf{1}\left(\xi_{[j]}\nu_{[j]} = 1\right)} \right), \tag{3.1}$$

where $\delta_{[i]}$, $\xi_{[i]}$ and $\nu_{[i]}$ are the concomitants of the ordered observed times $T_{(1)} \leq \cdots \leq T_{(n)}$, and $B_{h[i]}\,(x)$ is defined in equation (1.3). Note that the estimator in (3.1) is $\widehat{S}_h^c\left(T_{(n)}^1 \mid x\right)$, with $\widehat{S}_h^c\,(t \mid x)$ the estimator of the survival function proposed in (2.16), and $T_{(n)}^1 = \max_{i:\delta_i = 1} T_i$ the largest uncensored observed lifetime.

Proposition 3.1 The estimator $1 - \widehat{p}_h^c\,(x)$ has the following basic properties.

1. When there are no censored observations known to be cured, i.e., $\xi_i \nu_i = 0$ for $i = 1, \ldots, n$, $1 - \widehat{p}_h^c\,(x)$ reduces to the XP estimator in (1.5).

2. In the specific case that some individuals are classified as cured when their survival time exceeds a known fixed cure threshold, $1 - \widehat{p}_h^c\,(x)$ reduces to the XP estimator.

3. When there is no censoring, all the cure status indicators ν_i are observed ($\xi_i = 1, i = 1, \ldots, n$). In this case, $1 - \widehat{p}_h^c\,(x)$ reduces to the NW estimator of the cure probability:
$$1 - \widehat{p}_h^{\mathrm{NW}}\,(x) = \sum_{i=1}^{n} B_{hi}\,(x)\,\mathbf{1}\,(\nu_i = 1) = \frac{\sum_{i=1}^{n} K_h\,(x - X_i)\,\nu_i}{\sum_{j=1}^{n} K_h\,(x - X_j)}. \tag{3.2}$$
It must be kept in mind that when there is no censoring, the XP estimator will be zero.

4. In an unconditional setting, the proposed estimator is
$$1 - \widehat{p}_n^c = \prod_{i=1}^{n} \left(1 - \frac{\delta_{[i]}}{n - i + 1 + \sum_{j=1}^{i-1} \mathbf{1}\left(\xi_{[j]}\nu_{[j]} = 1\right)} \right). \tag{3.3}$$

In the particular case where an individual is known to be cured only if the observed time is greater than a known fixed time, $1-\widehat{p}_n^c$ reduces to the generalized maximum likelihood estimator in Laska and Meisner (1992):

$$1-\widehat{p}_n=\prod_{i=1}^{n}\left(1-\frac{\delta_{[i]}}{n-i+1}\right). \tag{3.4}$$

Moreover, if there are no individuals known to be cured, then $1-\widehat{p}_n^c$ becomes the unconditional version of the XP estimator.

The proof of these properties is outlined in Appendix B.

Proposition 3.2 The estimator $1-\widehat{p}_h^c(x)$ in (3.1) is the nonparametric local maximum likelihood estimator of $1-p(x)$.

The proof of Proposition 3.2 is given in Appendix B.

3.2.1 Asymptotic results

In this section, the asymptotic properties of $1-\widehat{p}_h^c(x)$ are studied. The required assumptions are listed in Section 2.2. Further define $\tau_H(x)=\inf\{t: H(t\mid x)=1\}$, $\tau_{S_0}(x)=\inf\{t: S_0(t\mid x)=0\}$ and $\tau_{G_0}(x)=\inf\{t: G_0(t\mid x)=1\}$. Note that $\tau_H(x)=\max\{\tau_{S_0}(x),\tau_{G_0}(x)\}$. Let $\tau_0=\sup_{x\in I}\tau_{S_0}(x)$, then it is required that

$$\tau_0<\tau_{G_0}(x) \text{ for any } x \text{ with probability 1.} \tag{3.5}$$

The condition (3.5) relies on the assumption that the follow-up is long enough for the support of the latency function $S_0(t\mid x)$ to be contained within the support of the distribution $G_0(t\mid x)$. This implies that all observations censored after the largest uncensored observed lifetime correspond to cured subjects, as the susceptible subjects will experience the event within the follow-up period. This condition guarantees that the proposed estimator does not overestimate the true probability of cure. A similar condition has been used in the related literature (Laska and Meisner, 1992; Xu and Peng, 2014; López-Cheda et al., 2017a,b). Xu and Peng (2014) pointed out that if $G_0(t\mid x)$ has a heavier tail than $S_0(t\mid x)$, then the condition (3.5) can be relaxed. Maller and Zhou (1994) proposed a test to assess whether a condition analogous to (3.5) is fulfilled in an unconditional setting. It is based on the difference between the largest observed time $T_{(n)}$ and the largest uncensored time $T^1_{(n)}$. If this interval is large, then there is sufficient follow-up time and (3.5) can be assumed. In the presence of covariates, one

may divide a given dataset into several subgroups according to the values of the covariates and apply this test in each subgroup.
The next theorem establishes an asymptotic representation for $1-\widehat{p}_h^c(x)$. The proof is included in Appendix B.

Theorem 3.1 (**Asymptotic representation**) Suppose that Assumptions $1-9$ and condition (3.5) hold, then for $x \in I$,

$$(1-\widehat{p}_h^c(x)) - (1-p(x)) = (1-p(x)) \sum_{i=1}^{n} \widetilde{B}_{hi}(x)\, \zeta\left(T_i, \delta_i, \xi_i, \nu_i, \tau_0, x\right) + R_n(x),$$

where $\zeta\left(T_i, \delta_i, \xi_i, \nu_i, t, x\right)$ is given in (2.21), $\widetilde{B}_{hi}(x)$ is defined in (2.22), and $R_n(x)$ satisfies

$$\sup_{x\in I} \mid R_n(x) \mid = O\left((\log n)^{3/4} (nh)^{-3/4}\right) \text{ a.s.} \tag{3.6}$$

The following corollary establishes the strong consistency of the estimator $1-\widehat{p}_h^c(x)$.

Corollary 3.1 (**Strong consistency**) Suppose that Assumptions $1-9$ and condition (3.5) hold. Then, for $x \in I$,

$$\sup_{x\in I} \mid \widehat{p}_h^c(x) - p(x) \mid = O\left((nh)^{-1/2} (\log n)^{1/2}\right) \quad \text{a.s.}$$

The corollary can be proved by considering the asymptotic representation in Theorem 3.1 and following similar arguments as those used in the proof of Corollary 2.1 for the estimator $1-\widehat{F}_h^c(t \mid x)$ when $t = T^1_{(n)}$. The next proposition, whose proof is in Appendix B, gives asymptotic expressions for the bias and variance of the estimator $1-\widehat{p}_h^c(x)$.

Proposition 3.3 (**Asymptotic bias and variance**) Suppose that Assumptions $1-9$ and condition (3.5) hold, then, the asymptotic bias and variance of the dominant term of $1-\widehat{p}_h^c(x)$ are, respectively,

$$\mu_{h,c}(x) = h^2 B_c(x) + O\left(h^4\right) \quad \text{and} \quad \sigma^2_{h,c}(x) = \frac{1}{nh} s_c^2(x) + O\left(\frac{h}{n}\right).$$

The function $B_c(x)$ in the dominant term of the bias is

$$B_c(x) = \left(c_{1,c}(x) + c_{2,c}(x)\right) d_K \tag{3.7}$$

with $d_K = \int v^2 K(v)dv$,

$$c_{1,c}(x) = \frac{2(1-p(x))' m'(x) + (1-p(x))'' m(x)}{2m(x)}, \tag{3.8}$$

and

$$c_{2,c}(x)=(1-p(x))\int_0^{\tau_0}\frac{G'(v^-\mid x)}{1-G(v^-\mid x)}\frac{d}{ds}\left(\frac{S'\left(s\mid x\right)}{S\left(s\mid x\right)}\right)\Big|_{s=v^-}dv. \tag{3.9}$$

Here $p'(x), p''(x), S'(t\mid x)$ and $G'(t\mid x)$ refer to the derivatives with respect to x. The function $s_c^2(x)$ in the dominant term of the variance is

$$s_c^2(x)=\frac{\left(1-p\left(x\right)\right)^2}{m\left(x\right)}\int_0^{\tau_0}\frac{dH^1\left(v^-\mid x\right)}{\left(1-H\left(v^-\mid x\right)+H^{11}(v^-\mid x)\right)^2}c_K, \tag{3.10}$$

with $c_K=\int K^2(v)dv$.

The following theorem establishes the asymptotic normality of $1-\widehat{p}_h^c\left(x\right)$. The proof is included in Appendix B.

Theorem 3.2 (**Asymptotic normality**) Suppose that Assumptions 1−9 and condition (3.5) are satisfied, then for $x\in I$ it follows that:

(i) If $nh^5\to 0$ and $(\log n)^3/(nh)\to 0$, then

$$(nh)^{1/2}\left(\widehat{p}_h^c(x)-p(x)\right)\xrightarrow{d}N(0,s_c^2(x)).$$

(ii) If $nh^5\to C$, where $C>0$ is a constant then

$$(nh)^{1/2}\left(\widehat{p}_h^c(x)-p(x)\right)\xrightarrow{d}N(C^{5/2}B_c(x),s_c^2(x)),$$

where $B_c(x)$ is defined in (3.7) and $s_c^2(x)$ in (3.10).

3.2.2 Effect of ignoring the cure status information

The use of the information given by the cure status has an impact on both the bias and variance of the proposed estimator of the cure probability $1-\widehat{p}_h^c(x)$. When the cure status is ignored in the estimation procedure and the observed times of the individuals known to be cured are considered as simple censored times, the asymptotic expressions of the bias and variance of the XP estimator of the cure rate, $1-\widehat{p}_h(x)$, are:

$$\mu_h(x)=h^2B(x)+O\left(h^4\right)\quad\text{and}\quad\sigma_h^2(x)=\frac{1}{nh}s^2(x)+O\left(\frac{h}{n}\right),$$

where $B(x)=\left(c_{1,c}(x)+c_2(x)\right)d_K$, with $c_{1,c}(x)$ in (3.8),

$$c_2(x)=(1-p(x))\int_0^{\tau_0}\frac{G_0'(v^-\mid x)}{1-G_0(v^-\mid x)}\frac{d}{ds}\left(\frac{S'\left(s\mid x\right)}{S\left(s\mid x\right)}\right)\Big|_{s=v^-}dv, \tag{3.11}$$

and

$$s^2(x)=\frac{\left(1-p\left(x\right)\right)^2}{m\left(x\right)}\int_0^{\tau_0}\frac{dH^1\left(v^-\mid x\right)}{\left(1-H\left(v^-\mid x\right)\right)^2}c_K.$$

The cure status information affects the bias of the proposed estimator only in the second term of $B_c(x)$ in (3.7). If the cure status information is ignored, the term $1-\pi(x)$ in $G(t \mid x)$ disappears and results to $c_1(x)$ in (3.11). Therefore, in terms of bias, the gain of knowing the cure status is not straightforward as it depends on the derivatives of $(1-\pi(x))$ and $G_0(t \mid x)$.

The effect of considering the cure status information on the variance is through the function $H^{11}(t \mid x)$ of $s_c^2(x)$ in (3.10). When the cure status information is ignored, then $H^{11}(t \mid x) = 0$ and, therefore, $s_c^2(x) \leq s^2(x)$ for all x. As a consequence, when the known cure status is taken into account for estimating the cure probability, the variance of the proposed estimator decreases asymptotically with respect to the XP estimator.

3.2.3 Bootstrap bandwidth selection

Here, a bootstrap bandwidth selector is proposed to choose the smoothing parameter h for the cure rate estimator $1-\widehat{p}_h^c(x)$. The principle is to select the bandwidth h that minimizes $\mathrm{MSE}_x^*(h)$, the bootstrap version of the MSE approximated by Monte Carlo as:

$$\mathrm{MSE}_x^*(h) \simeq \frac{1}{B}\sum_{b=1}^{B}(\widehat{p}_h^{c,*b}(x) - \widehat{p}_g^c(x))^2, \tag{3.12}$$

where $1-\widehat{p}_h^{c,*b}(x)$ is the proposed estimator computed with the bth bootstrap resample and bandwidth h. In addition, $1-\widehat{p}_g^c(x)$ is the proposed estimator computed with the original sample and a given pilot bandwidth g. The algorithm to compute the bootstrap bandwidth for a fixed covariate value x is as follows:

Step 1. With the original sample and the pilot bandwidth g, compute $1-\widehat{p}_g^c(x)$ in (3.1).

Step 2. Choose a dense enough grid of L bandwidths $\{h_1, \ldots, h_L\}$.

Step 3. Generate B bootstrap resamples $\{(X_i^{*(b)}, T_i^{*(b)}, \delta_i^{*(b)}, \xi_i^{*(b)}\nu_i^{*(b)}) : i = 1, \ldots, n\}$, for $b = 1, \ldots, B$.

Step 4. With the bth bootstrap resample and the bandwidth h_l compute $1-\widehat{p}_{h_l}^{c,*b}(x)$, for $l = 1, \ldots, L$,

Step 5. For $h_l, l = 1, \ldots, L$, compute the Monte Carlo approximation $\mathrm{MSE}_x^*(h_l)$ given by (3.12).

Step 6. The bootstrap bandwidth, h_x^*, is the bandwidth of the grid $\{h_1, \ldots, h_L\}$ that minimizes $\mathrm{MSE}_x^*(h)$ in (3.12).

The bootstrap resamples in Step 3 are generated following any of the two equivalent resampling algorithms introduced in Section 2.3.3. For computational efficiency (see López-Cheda et al., 2017a,b), we fixed $X_i^* = X_i$ instead of resampling it randomly from $\{X_1, \ldots, X_n\}$. It is suggested to use (2.38) for the pilot bandwidth g.

3.3 Alternative estimators of the cure rate

In this section we introduce some alternative estimators for the estimation of the cure rate with covariates. These estimators derive from the extension of unconditional estimators of the cure probability in the literature to the context with a continuous covariate X.

3.3.1 Competing risks estimators

The competing risks model considers that an individual is exposed to J types of failure or competing risks. For $j \in \{1, \ldots, J\}$, let Y_j the time until the failure of type j happens, and consider the random pair (Y_F, D), where $Y_F = \min(Y_1, \ldots, Y_J)$ is a non-negative random variable representing the time until the first failure, and D takes a value from the set $\{1, 2, \ldots, J\}$ to indicate the type of failure. Let C be a censoring variable. Under right random censoring, the observations (Y_F, D) will be incomplete if follow-up ends before any failure occurs. In this situation only (T, Δ) is observed, where $T = \min(Y_F, C) = \min(Y_1, \ldots, Y_J, C)$ is the possibly censored observed time, and $\Delta = \mathbf{1}(Y_F < C)D$ is the type of event in the case a terminal event occurs and $\Delta = 0$ indicates that the failure type is unknown and the failure time is right-censored. The censoring mechanism is assumed to be non-informative (Lagakos, 1979). This competing risks model assumes that an individual will fail from a particular risk $j \in \{1, \ldots, J\}$, chosen by a stochastic mechanism at the outset. This general competing risks model usually assumes that all patients will eventually experience one of the J possible types of risks if there is sufficient follow-up and, therefore, do not consider the possibility of cure.

In the MCM with cured individuals randomly observed, Betensky and Schoenfeld

(2001) stated that the event of interest and cure can be regarded as two competing risks, in which cures are random and only the minimum between the cure and event times is observed. The probability of cure is then simply the cumulative incidence function of the cure evaluated at the largest observed time, or just 1 minus the cumulative incidence function of the event of interest.

In this section, we adopt this perspective and introduce a competing risks model for the MCM when the cure status is partially observed in the presence of covariates. Here, as in Betensky and Schoenfeld (2001), the observed times of the individuals known to be cured are considered as a competing risks for the event of interest. Let $\{Y_E, Y_c\}$ be the latent failure times of 2 type failures: the event of interest (E) and the classification of an individual as cured (c). Let $Y_F = \min(Y_E, Y_c)$ be the time of the first failure and C the censoring time. For right censored competing risks data, let $T = \min(Y_E, Y_c, C)$ be the observed time, and the uncensoring indicator $\Delta = \mathbf{1}(Y_F < C)\, D$ where $D \in \{1, 2\}$ is the type of risk. In this context, the observed sample in the MCM with the cure status partially known $\{(T_i, \delta_i, \xi_i \nu_i), i = 1, \ldots, n\}$ can be written as $\{(T_i, \Delta_i), i = 1, \ldots, n\}$, where

$$\Delta_i = \begin{cases} 0 & \text{if } \delta_i = 0, \xi_i\nu_i = 0 \text{ (censored)} \\ 1 & \text{if } \delta_i = 1, \xi_i\nu_i = 0 \text{ (event observed)} \\ 2 & \text{if } \delta_i = 0, \xi_i\nu_i = 1 \text{ (known to be cured)}. \end{cases}$$

The cumulative incidence function (CIF) of the event of interest E is the probability that a failure of type 1 occurs at or before time t:

$$F_1(t \mid x) = P(Y_F \leq t, D = 1 \mid X = x).$$

The CIF of the second competing risk (individual known to be cured) is the probability that a failure of type 2 occurs at or before time t:

$$F_2(t \mid x) = P(Y_F \leq t, D = 2 \mid X = x).$$

The probability of cure is then simply the cumulative incidence function of the competing risk *cure* (c) evaluated at infinity or the complementary of the cumulative incidence function of the event of interest (E) evaluated at infinity:

$$1 - p(x) \quad = \quad P(Y_E = \infty \mid X = x) = 1 - \lim_{t \to \infty} F_1(t \mid x).$$

Equivalently,

$$1 - p(x) \quad = \quad P(Y_c < \infty \mid X = x) = \lim_{t \to \infty} F_2(t \mid x).$$

The conditional version of the estimators of the CIFs in Klein and Moeschberger

(2003) are the following (see Effraimidis and Dahl, 2014):

$$\begin{aligned}
\widehat{F}_{1,h}(t \mid x) &= \sum_{i=1}^{n} \frac{\delta_{[i]} B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x)} \prod_{k=1}^{i-1} \left(1 - \frac{\left(\delta_{[k]} + \xi_{[k]}\nu_{[k]}\right) B_{h[k]}(x)}{\sum_{j=k}^{n} B_{h[j]}(x)}\right) \\
&= \sum_{i=1}^{n} \frac{\delta_{[i]} B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x)} \widehat{S}_h\left(T_{(i)}^{-} \mid x\right), \qquad (3.13) \\
\widehat{F}_{2,h}(t \mid x) &= \sum_{i=1}^{n} \frac{\xi_{[i]}\nu_{[i]} B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x)} \prod_{k=1}^{i-1} \left(1 - \frac{\left(\delta_{[k]} + \xi_{[k]}\nu_{[k]}\right) B_{h[k]}(x)}{\sum_{j=k}^{n} B_{h[j]}(x)}\right) \\
&= \sum_{i=1}^{n} \frac{\xi_{[i]}\nu_{[i]} B_{h[i]}(x) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}(x)} \widehat{S}_h\left(T_{(i)}^{-} \mid x\right), \qquad (3.14)
\end{aligned}$$

where $\widehat{S}_h(t \mid x)$ is the Beran estimator obtained by treating any of the competing risks as an event. It may be shown that the sum of the cumulative incidences for all competing risks, $\widehat{F}_{1,h}(t \mid x)$ and $\widehat{F}_{2,h}(t \mid x)$, is $\widehat{F}_h(t \mid x) = 1 - \widehat{S}_h(t \mid x)$ where $\widehat{S}_h(t \mid x)$ is the aforementioned Beran's estimator.

Proposition 3.4 The estimation of the conditional CIFs given in (3.13) and (3.14) allows us to model the conditional cure probability. Thus, the conditional probability of cure $1 - p(x)$ can be estimated by

$$\begin{aligned}
1 - \widehat{p}_{1,h}(x) &= 1 - \lim_{t \to \infty} \widehat{F}_{1,h}(t \mid x) \\
&= 1 - \sum_{i=1}^{n} \frac{\delta_{[i]} B_{h[i]}(x)}{\sum_{j=i}^{n} B_{h[j]}(x)} \widehat{S}_h\left(T_{(i)}^{-} \mid x\right), \qquad (3.15) \\
1 - \widehat{p}_{2,h}(x) &= \lim_{t \to \infty} \widehat{F}_{2,h}(\ t \mid x) = \sum_{i=1}^{n} \frac{\xi_{[i]}\nu_{[i]} B_{h[i]}(x)}{\sum_{j=i}^{n} B_{h[j]}(x)} \widehat{S}_h\left(T_{(i)}^{-} \mid x\right). \qquad (3.16)
\end{aligned}$$

If the last observation is an event or an observed cured individual, then $1 - \widehat{p}_{1,h}(x) = 1 - \widehat{p}_{2,h}(x)$. If, however, the last observation is censored, $1 - \widehat{p}_{1,h}(x)$ and $1 - \widehat{p}_{2,h}(x)$ are not equivalent. In this case, $1 - \widehat{p}_{1,h}(x)$ is an upper bound for the cure rate $1 - p(x)$ and $1 - \widehat{p}_{2,h}(x)$ is a lower bound. Note that, in the absence of censoring,

$$1 - \frac{\left(\delta_{[k]} + \xi_{[k]}\nu_{[k]}\right) B_{h[k]}(x)}{\sum_{j=k}^{n} B_{h[j]}(x)} = 1 - \frac{B_{h[k]}(x)}{\sum_{j=k}^{n} B_{h[j]}(x)} = \frac{\sum_{j=k+1}^{n} B_{h[j]}(x)}{\sum_{j=k}^{n} B_{h[j]}(x)}$$

and therefore

$$\begin{aligned}
\widehat{S}_h\left(T_{(i)}^{-} \mid x\right) &= \prod_{k=1}^{i-1} \left(1 - \frac{\left(\delta_{[k]} + \xi_{[k]}\nu_{[k]}\right) B_{h[k]}(x)}{\sum_{j=k}^{n} B_{h[j]}(x)}\right) = \prod_{k=1}^{i-1} \frac{\sum_{j=k+1}^{n} B_{h[j]}(x)}{\sum_{j=k}^{n} B_{h[j]}(x)} \\
&= \sum_{j=i}^{n} B_{h[j]}(x).
\end{aligned}$$

So the cure probability estimator is

$$\begin{aligned}
1-\widehat{p}_{2,h}\left(x\right) &= 1-\sum_{i=1}^{n}\frac{\delta_{[i]}B_{h[i]}\left(x\right)}{\sum_{j=i}^{n}B_{h[j]}\left(x\right)}\widehat{S}_{h}\left(T_{(i)}^{-}\mid x\right)\\
&= 1-\sum_{i=1}^{n}\frac{\delta_{[i]}B_{h[i]}\left(x\right)}{\sum_{j=i}^{n}B_{h[j]}\left(x\right)}\sum_{j=i}^{n}B_{h[j]}\left(x\right)=1-\sum_{i=1}^{n}\delta_{[i]}B_{h[i]}\left(x\right)\\
&= \sum_{i=1}^{n}\xi_{[i]}\nu_{[i]}B_{h[i]}\left(x\right).
\end{aligned}$$

Similarly,

$$1-\widehat{p}_{2,h}\left(x\right)=\sum_{i=1}^{n}\xi_{[i]}\nu_{[i]}B_{h[i]}\left(x\right).$$

That is, the cure rate estimators reduce to the sum of the weights of the individuals known to be cured.

This approach to the estimation of the cure rate can be viewed as arising from a redistribution to the right algorithm (Efron, 1967). In particular, the mass of $B_{hi}\left(x\right)$ initially assigned to the censored observations (neither event nor observed to be cured) is redistributed equally to all subjects at risk for the event and cure at the time of censoring. The cure rate is then simply the weighted sum of the mass attached to each subject that is cured.

3.3.2 Multiply imputed NW estimator

The proposed estimator in (3.1) is based on the relationship between the cure probability $1-p(x)$ and the survival function $S(t\mid x)$. So, to estimate the probability of not experiencing the event it requires the observations $\{(T_i,\delta_i), i=1,\ldots,n\}$. Nonetheless, the cure probability can also be written as $1-p(x)=E(\nu\mid X=x)$, i.e., the conditional expectation of the cure status ν, or equivalently $1-p=E(\nu)$ for an unconditional setting. An estimator based on this latter relationship would only require the observed values of the covariate X and the cure status ν, dismissing the observed values of (T,δ).

The NW estimator is one of the most frequently used estimators in nonparametric regression. So, the estimator in (3.2) might be considered for the estimation of the cure probability $1-p(x)=E(\nu\mid X=x)$. Similarly, the unconditional cure probability $1-p=E(\nu)$ might easily be estimated using the empirical estimator $1-\widehat{p}=\sum_{i=1}^{n}\nu_i/n$. These methods require that the cure status ν is completely observed. However, in the present setup, the cure status ν remains unknown for some of the censored observations. There has been extensive work dealing with estimating the unconditional and conditional mean in a regression setting when the response variable is only partially observed (Hsu et al., 2016; Verhasselt et al., 2019; Vakulenko-Lagun et al., 2020).

Aerts et al. (2002) developed a fully nonparametric local multiple imputation (MI) procedure to estimate the unconditional mean of a variable in the presence of missing response data. When the cure status is not completely observed because of censoring but it is partially available, their methodology can be applied to the estimation of $1 - p$. To the best of our knowledge, the MI methodology in Aerts et al. (2002) has not been extended to estimate the conditional mean. In this section an estimator for the cure probability in the presence of a covariate $1 - p(x)$ is proposed.

It is important to define the nature of the missingness mechanism, as it highly influences the performance of statistical techniques that deal with missing data. The MI estimator (Aerts et al., 2002) requires the strongly ignorable missing at random (siMAR) assumption (Rosenbaum and Rubin, 1983), which implies that given ν and X, the probability that the cure status is observed depends only on the covariate X but not on the response variable ν:

$$E(\xi \mid X) = E(\xi \mid X, \nu). \tag{3.17}$$

This is weaker than missingness completely at random (MCAR) since dependence on the observed variable X is allowed. It is important to note that this siMAR condition is not fulfilled under the MCM model with the cure status partially known if Y and C are conditionally independent given $X = x$, as the probability of observing the cure status is different for the cured ($\nu = 1$) and the susceptible ($\nu = 0$) individuals, and therefore it depends on the cure status:

$$\begin{aligned}
E(\xi \mid X, \nu = 1) =& P(\xi = 1 \mid X, Y = \infty) = P(C = \infty \mid X, Y = \infty) = \pi(X),\\
E(\xi \mid X, \nu = 0) =& P(\xi = 1 \mid X, Y < \infty) = P(Y < C \mid X, Y < \infty)\\
=& P(Y < C, C < \infty \mid X, Y < \infty) + P(C = \infty \mid X, Y < \infty)\\
=& P(Y < C \mid X, Y < \infty, C < \infty)(1 - \pi(X)) + \pi(X).
\end{aligned}$$

Unless $P(Y < C \mid X, Y < \infty, C < \infty) = 0$, which yields the time of all the susceptible individuals to be censored, the siMAR condition in (3.17) cannot be assumed if Y and C are independent conditionally on $X = x$. Nonetheless, note that the higher the value of $\pi(X) = P(C = \infty \mid X)$, the smaller the difference between $E(\xi \mid X, \nu = 1)$ and $E(\xi \mid X, \nu = 0)$, and the closer the siMAR assumption to hold.

The main idea in the approach of Aerts et al. (2002) is to use the assumed regression relationship between X and ν to impute locally the missing observations of ν. This idea is extended to estimate the conditional expectation for a continuous covariate X. An outline of the algorithm is:

Step 1. (Resampling step) Fix an integer M, for $m = 1, \dots, M$ perform a nonparametric resampling of the observed data. That is, for each observation $i = 1, \dots, n$, if the cure status is observed ($\xi_i = 1$) generate $\nu_i^{*(m)}$ from the distribution $\mathcal{L}(X_i)$ with cumulative distribution function

$$\sum_{j=1}^{n} B_{g_1 j}^{\xi}(X_i) \mathbf{1}(\nu_j \leq u)$$

where $B_{g_1 j}^{\xi}(x)$ are the kernel weights with bandwidth g_1:

$$B_{g_1 j}^{\xi}(x) = \frac{\xi_j K_{g_1}(x - X_j)}{\sum_{i=1}^{n} \xi_i K_{g_1}(x - X_i)}.$$

Step 2. (Imputation step) Given the resampled data from Step 1., the missing values of ν are imputed using local resampling. More specifically, conditionally on the resampled data $\{(X_i, \nu_i^{*(m)}, \xi_i) : i = 1, \dots, n\}$, a second distribution $\mathcal{L}^*(X_i)$ is constructed, with cumulative distribution function

$$\sum_{j=1}^{n} B_{g_2 j}^{\xi}(X_i) \mathbf{1}\left(\nu_j^{*(m)} \leq u\right)$$

where the kernel weights $B_{g_2 j}^{\xi}(x)$ are computed with a second bandwidth g_2. Then, if ν_i is missing, generate $\nu_i^{+,m}$ from $\mathcal{L}^*(X_i)$.

Step 3. (Computation of the final estimator) For $\widetilde{\nu}_i^m = \xi_i \nu_i + (1 - \xi_i)\nu_i^{+,m}$, let $1 - \widehat{p}_n^m = (1/n) \sum_{i=1}^{n} \widetilde{\nu}_i^m$ be the empirical estimator of the cure probability with the mth augmented dataset. The multiple imputation (MI) estimator for the cure probability $1 - p$ is

$$1 - \widehat{p}_n^{\mathrm{MI}} = \frac{1}{M} \sum_{m=1}^{M} (1 - \widehat{p}_n^m). \tag{3.18}$$

Analogously, let us define $1 - \widehat{p}_h^m(x) = \sum_{i=1}^{n} B_{hi}(x) \widetilde{\nu}_i^m$ as the NW estimator in (3.2) computed with bandwidth h and the mth augmented dataset. The final multiply imputed NW (MI-NW) estimator for the cure probability $1 - p(x)$ is

$$1 - \widehat{p}_h^{\text{MI-NW}}(x) = \frac{1}{M} \sum_{m=1}^{M} (1 - \widehat{p}_h^m(x)). \tag{3.19}$$

Note that Step 1 is needed to fully account for all uncertainty in predicting the missing values by adding extra variability into the multiply imputed values (Efron, 1994). Under conditions similar to those in Cheng (1994), Aerts et al. (2002) showed that the proposed estimator of $1-p$ in (3.18) is consistent, and provided asymptotic expressions for the bias and variance. Next, the asymptotic expressions of the bias and variance for the MI-NW estimator in (3.19), following the ideas in Aerts et al. (2002) are derived. The proof is deferred to Appendix B.

Proposition 3.5 Suppose that the siMAR condition and Assumptions 1 (i), 2 (i), 3 (i), 8 and 10 hold. Also, the bandwidths h, g_1, g_2 satisfy $h \to 0$, $g_1 \to 0$, $g_2 \to 0$, $nh \to \infty$, $ng_1 \to \infty$ and $ng_2 \to \infty$ as $n \to \infty$. The asymptotic bias of $1-\widehat{p}_h^{\text{MI-NW}}(x)$ is

$$\mu_{g_1,g_2,h}^{\text{MI-NW}}(x) = h^2 c_{1,c}(x) + \left(g_1^2 + g_2^2\right) c_{2,\text{MI-NW}}(x) + o\left((h^2 + g_1^2 + g_2^2)^2\right),$$

where $c_{1,c}(x)$ is defined in (3.8), and

$$c_{2,\text{MI-NW}}(x) = \frac{(1-\pi(x))\left[\pi(x)(1-p(x))m(x)\right]''}{2m(x)\pi(x)} d_K. \tag{3.20}$$

If the bandwidths are $g_1/h \to C_1$ and $g_2/h \to C_2$, then the asymptotic variance is

$$\begin{aligned}\sigma_{h,\text{MI-NW}}^2(x) =& \frac{1}{nh}\frac{1-p(x)}{m(x)}\Bigg(\frac{c_K(1-\pi(x))p(x)}{M\pi(x)} \\ &+ \Bigg\{\pi(x)c_K + (1-\pi(x))\Bigg[c_{K,C_1,C_2} + \frac{1-\pi(x)}{\pi(x)} d_{K,C_1,C_2} \\ &+ (1-p(x))\left(c_K + 2c_{K,C_2} + \frac{1-\pi(x)}{\pi(x)}(c_{K,C_1,C_2} + 2d_{K,C_1,C_2})\right)\Bigg]\Bigg\}\Bigg) \\ &+ \frac{2}{ng_1}(1-p(x))^2 \frac{1-\pi(x)}{\pi(x)} K(0) + o\left((Mnh)^{-1}\right) + o\left((nh)^{-1}\right) \\ &+ o\left((ng_1)^{-1}\right),\end{aligned} \tag{3.21}$$

where $c_{K,C} = \iint K(u)K(v)K(u+Cv)dudv$,

$$c_{K,C_1,C_2} = \iiint K(u)K(v)K(w)K(u+C_1v+C_2w)dudvdw$$

and

$$d_{K,C_1,C_2} = \iiint K(u)K(v)K(w)K(u+C_1v+C_2(u+w))dudvdw.$$

The term $h^2c_{1,c}(x)$ in the bias, which also appears in the bias $\mu_{h,c}(x)$ of the proposed estimator in (3.8) and in the bias $\mu_h(x)$ of the XP estimator, is the dominant term of the bias of the NW estimator of $1-p(x)$, while $(g_1^2+g_2^2)\,c_{2,\text{MI-NW}}(x)$, the second term in the bias, stems from the multiple imputation procedure in Steps 1

and 2 above. The comparison in terms of bias of the proposed estimator of the cure probability in Section 3.2 and the MI-NW estimator is a trade-off between the terms $c_{2,c}(x)$ in (3.9) and $c_{2,\text{MI-NW}}(x)$ in (3.20).
As for the variance, note that if $C_1 = C_2 = 0$ then $c_{K,C_2} = c_{K,C_1,C_2} = d_{K,C_1,C_2} = c_K$, whereas if $C_1 = \infty$ or $C_2 = \infty$, then $c_{K,C_2} = c_{K,C_1,C_2} = d_{K,C_1,C_2} = 0$. It should be noted that the comparison, in terms of variance, between the proposed estimator and the MI-NW estimator is not straightforward. It is easy to prove that in the case of no missingness, the dominant term of the bias reduces to that of the NW estimator $c_{1,c}(x)$, whereas the leading term of the variance becomes $(1/nh)(\sigma^2(x)+\mu^2(x))/m(x)$, where $\sigma^2(x) = \text{Var}(\nu \mid X = x) = p(x)(1-p(x))$ and $\mu(x) = E(\nu \mid X = x) = 1 - p(x)$.

3.4 Simulation study

A simulation study was conducted to assess the finite sample performance of the proposed estimator, $1 - \widehat{p}_h^c(x)$. The estimator $1 - \widehat{p}_h^c(x)$ is compared with:

(a) the competing risks estimators $1 - \widehat{p}_{1,h}(x)$ (CR1) in (3.15) and $1 - \widehat{p}_{2,h}(x)$ (CR2) in (3.16),

(b) the XP estimator $1 - \widehat{p}_h(x)$ in (1.5), which does not include the cure status information,

(c) the MI-NW estimator $1 - \widehat{p}_h^{\text{MI-NW}}(x)$ in (3.19) with $M = 5$ multiple imputations, which requires siMAR assumption (untrue in the MCM if Y and C are conditionally independent given $X = x$),

(d) the semiparametric estimator $1 - p(x; \widehat{\boldsymbol{\gamma}})$ by Bernhardt (2016), which considers a logistic regression model to fit the probability of cure with an EM algorithm for estimating the regression parameter $\boldsymbol{\gamma}$.

Data were generated from the MCM, where the latency part was modeled using the truncated exponential distribution in (2.40). Six different scenarios characterized by the cure probability function, $1 - p(x)$, were considered. As can be seen in Table 3.1, the cure probability displays a wide range of functions forms. The proportion of individuals identified as being cured was set to $\pi(x) = 0.2$ and 0.8. The censoring time C was generated, independently of X and Y, so that $C = \infty$ with probability $\pi(x)$, and with probability $1 - \pi(x)$, C was generated

Table 3.1: Characteristics of the simulated scenarios.

Scenario		$1-p(x)$	**% censoring**		**% cured**
			$\pi(x)=0.2$	$\pi(x)=0.8$	
1	Logistic function	$(1+\exp(0.476+0.358x))^{-1}$	48.0	47.0	46.0
2	Cubic function	$0.5-x^3/16000$	52.0	51.0	50.0
3	Linear function	$0.5-0.025x$	51.0	50.4	50.0
4	Low constant	0.2	22.6	20.5	20.0
5	High constant	0.8	80.8	80.3	80.0
6	Convex function	$0.0025x^2$	36.0	34.0	33.3

from a Weibull distribution with shape parameter $\alpha = 2$, scale parameter $\beta = 2$, and density function

$$g(t;\beta,\alpha) = \beta\alpha^{-\beta}t^{\beta-1}\exp\left(-t/\alpha\right)^{\beta}.$$

The covariate X was uniformly distributed on the interval $[-20, 20]$. Note that $S_0(t \mid x)$ is truncated at $\tau_0 = 4.605$, so that the support for C is larger than the support of Y in order to fulfill condition (3.5). Depending on the scenario, the percentage of censored observations ranged from 22.6% (in Scenario 4 with $\pi(x) = 0.2$) to 80.8% (in Scenario 5 with $\pi(x) = 0.2$). For each scenario, 1000 datasets of sample sizes $n = 50, 100$ and 200 were generated.

Two different designs were considered. They differ with respect to the distribution of the observed times of the individuals known to be cured, represented by $H^{11}(t \mid x)$. In the first design, Design 1, the observed lifetimes of the individuals known to be cured were simulated to be falling within the largest censored times. Under this setup, no big differences are expected between the proposed survival estimator (equivalent to Beran's estimator computed with the observed cure times shifted to be arbitrarily large time) or ignoring the known cure status (Beran's estimator computed with the unmodified observed times). This design was intended to reflect the pattern of the observed lifetimes of the patients known to be cured in the breast cancer data. In the second design, Design 2, the distribution of the observed times of the known cured patients in COVID-19 data is mimicked. In this case the known cured observations were simply chosen at random among the censored observations. Large differences are now expected between using the available cure status with the estimator $1-\widehat{p}_h^{\mathrm{c}}(x)$ and ignoring that information with XP estimator $1-\widehat{p}_h(x)$.

The first goal was to evaluate the small sample size performance of $1-\widehat{p}_h^{\mathrm{c}}(x)$ in terms of the squared bias, variance and MSE when the optimal bandwidth is used. For all the nonparametric estimators, the search for the optimal bandwidth h was performed in a grid of 21 values ranging from 1.5 to 100 and equispaced on a logarithmic scale. Besides, the pilot bandwidths required by the MI-NW estimator for the local resampling step (g_1) and the imputation step (g_2) were searched in a grid of 11 bandwidths equispaced from 1.5 to 100 on a logarithmic scale. The Epanechnikov kernel was used.

The MSE of $1-\widehat{p}_h^{\mathrm{c}}(x)$, $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$, $1-\widehat{p}_h(x)$ and $1-\widehat{p}_h^{\text{MI-NW}}(x)$, all of them computed with the corresponding optimal bandwidths, and the MSE of $1-p(x;\widehat{\boldsymbol{\gamma}})$ when $n=100$, $\pi(x)=0.8$ for Design 1 are illustrated in Figure 3.1. Summarizing all the scenarios, in general the proposed estimator $1-\widehat{p}_h^{\mathrm{c}}(x)$ has smaller MSE than XP estimator $1-\widehat{p}_h(x)$ for most values of X. This shows the loss of efficiency incurred in if the known cures are not incorporated in the estimation methodology. As expected, in Scenario 1, the semiparametric estimator behaves well since it fits a logistic regression for the cure probability. However, the estimator $1-\widehat{p}_h^{\mathrm{c}}(x)$ is competitive for a wide range of values close to $x=-20$ and $x=20$, and even beats $1-p(x;\widehat{\boldsymbol{\gamma}})$ for some values of the covariate around $x=0$. The estimator $1-\widehat{p}_h^{\mathrm{c}}(x)$ outperforms $1-p(x;\widehat{\boldsymbol{\gamma}})$ in Scenarios 2–6, where the underlying logistic model assumption for the cure probability in $1-p(x;\widehat{\boldsymbol{\gamma}})$ is not met. Finally, it must be noted that $1-\widehat{p}_h^{\mathrm{c}}(x)$ is quite competitive with respect to $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$ and $1-\widehat{p}_h^{\text{MI-NW}}(x)$, showing in general a better behavior.

The MSE results obtained for Design 2 are presented in Figure 3.2. Note that while the performance of $1-\widehat{p}_h^{\mathrm{c}}(x)$, $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$, $1-\widehat{p}_h^{\text{MI-NW}}(x)$ and $1-p(x;\boldsymbol{\gamma})$ is affected by the design, that of $1-\widehat{p}_h(x)$ is not affected as it ignores the information provided by the observations identified as cured. As in Design 1, $1-\widehat{p}_h^{\mathrm{c}}(x)$ outdoes $1-\widehat{p}_h(x)$ for most values of X. The differences in the squared bias between $1-\widehat{p}_h^{\mathrm{c}}(x)$ and the competing estimators are quite apparent in Scenarios 1–3 and 6. As it can be seen, in all the scenarios $1-\widehat{p}_h^{\mathrm{c}}(x)$ outperforms $1-\widehat{p}_{1,h}(x)$ and $1-\widehat{p}_{2,h}(x)$. Table 3.2 collects the MSE, squared bias and variance of all the estimators. The estimators $1-\widehat{p}_h^{\mathrm{c}}(x)$ and $1-\widehat{p}_h^{\text{MI-NW}}(x)$ are competing with each other with the overall better performance reflected in $1-\widehat{p}_h^{\mathrm{c}}(x)$.

When $n=200$ in Design 2, it can be seen that the differences in squared bias are much smaller for $1-\widehat{p}_h^{\mathrm{c}}(x)$, $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$, $1-\widehat{p}_h^{\text{MI-NW}}(x)$ and $1-p(x;\widehat{\boldsymbol{\gamma}})$, see Figure 3.3 (top) and Table 3.3. Regarding the variance, $1-\widehat{p}_h^{\mathrm{c}}(x)$ performs better in most scenarios.

Figure 3.3 (bottom) and Table 3.4 show the simulation results when $n = 100$ and $\pi(x) = 0.2$. Interestingly, when $\pi(x) = 0.2$, $1 - \widehat{p}_h^c(x)$ is still efficient and even beats $1 - \widehat{p}_h(x)$ for most values of X. Besides, $1 - \widehat{p}_h^{\text{MI-NW}}(x)$ and $1 - \widehat{p}_{2,h}(x)$ perform poorly due to a significant increase in both squared bias and variance, which leads to poor MSE results. This suggests that there is an advantage in applying $1-\widehat{p}_h^c(x)$ even when one has a few individuals identified as being cured. A simulation study was conducted to evaluate the practical performance of the bandwidth selector discussed in Section 3.2.3, using $B = 1000$ resamples and a grid of bandwidths from 1.5 to 100. Figure 3.4 shows the quartile of the selected bootstrap bandwidth h_x^* for Scenarios $1-6$ under Design 2. The optimal bandwidth was also computed and was compared to h_x^*. The performance of h_x^* varies depending on the scenario, but in general it seems to perform well in all scenarios. The choice of the bandwidth seems to be more important in Scenarios $1-3$ and 6, as different bandwidths result in slightly different MSE. In Scenarios 4 and 5, different bandwidths yield approximately the same MSE. In this case, the bootstrap bandwidth being relatively far from the optimal bandwidth does not entail a significant loss of efficiency, see Figure 3.5.

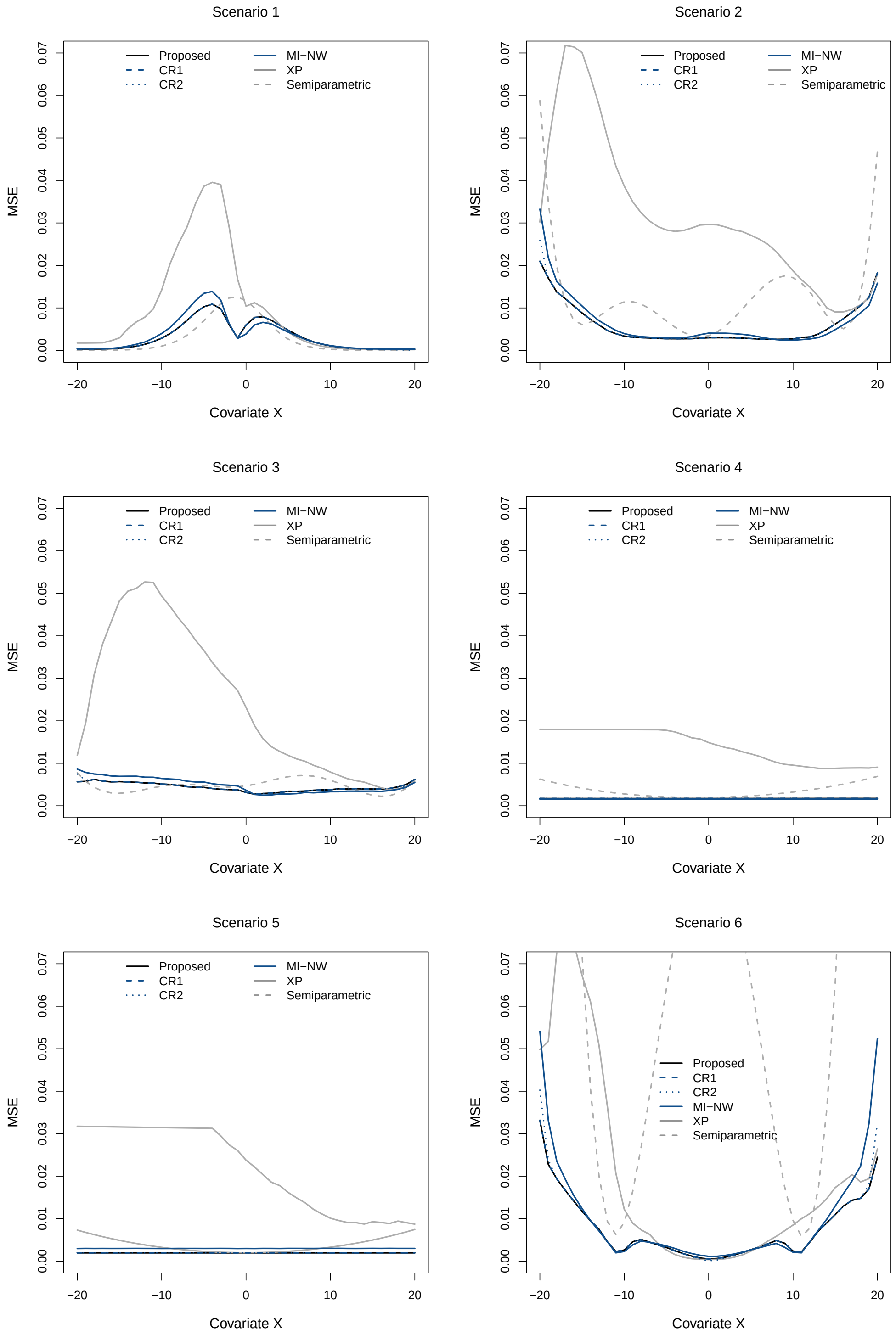

Figure 3.1: MSE of $1-\widehat{p}_h^c(x)$, $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$, $1-\widehat{p}_h^{\text{MI-NW}}(x)$, $1-\widehat{p}_h(x)$ (all of them computed with the optimal bandwidth), and $1-p(x;\widehat{\boldsymbol{\gamma}})$ in the simulated scenarios and under Design 1, for $\pi(x)=0.8$ and $n=100$.

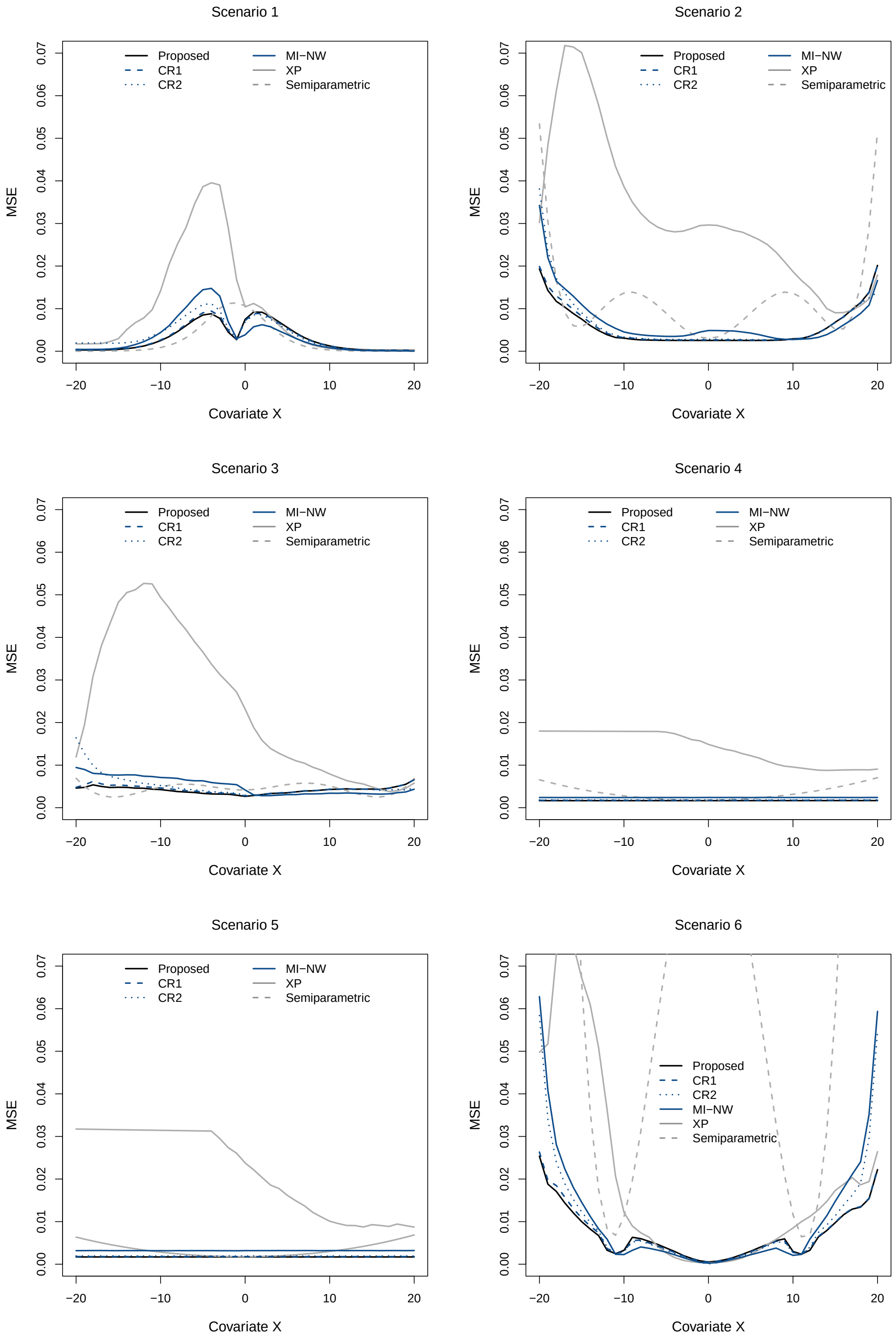

Figure 3.2: MSE of $1-\widehat{p}_h^c(x)$, $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$, $1-\widehat{p}_h^{\text{MI-NW}}(x)$, $1-\widehat{p}_h(x)$ (all computed with the optimal bandwidth), and $1-p(x;\widehat{\gamma})$ in the simulated scenarios and under Designs 2 for $\pi(x)=0.8$ and $n=100$.

Table 3.2: Squared bias (Bias^2), variance (Var) and MSE of $1-\widehat{p}_h^c(x)$, $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$, $1-\widehat{p}_h(x)$, $1-\widehat{p}_h^{\text{MI-NW}}(x)$, all computed with the optimal MSE bandwidth, and $1-p(x;\widehat{\gamma})$ in the simulated scenarios and under Designs 1 and 2, for $\pi(x)=0.8$ and $n=100$.

		$1-\widehat{p}_h^c(x)$				$1-\widehat{p}_{1,h}(x)$				$1-\widehat{p}_{2,h}(x)$				$1-\widehat{p}_h(x)$				$1-\widehat{p}_h^{\text{MI-NW}}(x)$				$1-p(x;\widehat{\gamma})$		
Scenario	x	h_x	**Bias**^{2} $\times 10^3$	**Var** $\times 10^3$	**MSE** $\times 10^3$	h_x	**Bias**^{2} $\times 10^3$	**Var** $\times 10^3$	**MSE** $\times 10^3$	h_x	**Bias**^{2} $\times 10^3$	**Var** $\times 10^3$	**MSE** $\times 10^3$	h_x	**Bias**^{2} $\times 10^3$	**Var** $\times 10^3$	**MSE** $\times 10^3$	h_x	**Bias**^{2} $\times 10^3$	**Var** $\times 10^3$	**MSE** $\times 10^3$	**Bias**^{2} $\times 10^3$	**Var** $\times 10^3$	**MSE** $\times 10^3$
											Design 1													
1	-10	5.288	0.302	2.579	2.882	5.288	0.302	2.579	2.882	5.288	0.302	2.579	2.882	2.283	0.317	13.885	14.202	2.283	1.101	2.873	3.974	0.001	0.998	0.999
	0	15.109	1.780	4.205	5.985	15.109	1.771	4.207	5.978	15.109	1.771	4.207	5.978	65.707	0.028	10.381	10.409	9.928	1.090	2.764	3.854	0.493	11.191	11.684
	10	6.523	0.098	0.966	1.064	6.523	0.095	0.957	1.052	6.523	0.081	0.938	1.019	8.047	0.003	0.678	0.681	2.816	0.157	0.964	1.121	0.000	0.253	0.253
2	-10	28.368	0.171	3.193	3.364	28.368	0.175	3.194	3.370	28.368	0.175	3.194	3.370	28.368	17.188	21.477	38.665	9.928	0.480	3.473	3.953	6.471	4.930	11.401
	0	81.060	0.367	2.612	2.978	81.060	0.372	2.613	2.985	81.060	0.372	2.613	2.985	100	14.321	15.326	29.647	43.174	1.623	2.437	4.060	0.343	3.108	3.451
	10	34.996	0.009	2.686	2.695	34.996	0.008	2.687	2.695	34.996	0.008	2.687	2.695	100	3.640	15.059	18.699	22.995	0.005	2.394	2.399	12.906	4.213	17.119
3	-10	12.247	0.539	4.571	5.111	12.247	0.543	4.576	5.119	12.247	0.543	4.576	5.119	12.247	15.379	33.983	49.362	5.288	1.581	4.843	6.424	0.559	4.055	4.613
	0	81.060	0.419	2.727	3.146	81.060	0.423	2.730	3.152	81.060	0.423	2.730	3.152	100	9.828	13.330	23.158	22.995	1.064	2.596	3.661	0.377	4.314	4.691
	10	15.109	0.094	3.692	3.785	15.109	0.091	3.695	3.786	15.109	0.091	3.695	3.786	18.640	0.369	7.561	7.930	6.523	0.003	3.315	3.318	2.419	3.598	6.017
4	-10	81.060	0.229	1.483	1.712	81.060	0.249	1.483	1.732	81.060	0.250	1.484	1.734	100	14.109	3.821	17.930	34.996	0.262	1.311	1.573	0.202	2.583	2.785
	0	100	0.231	1.486	1.717	100	0.250	1.485	1.736	100	0.252	1.487	1.738	9.928	8.612	6.260	14.872	43.174	0.270	1.309	1.579	0.343	1.606	1.950
	10	100	0.232	1.491	1.723	100	0.251	1.490	1.742	100	0.252	1.492	1.744	12.247	4.237	5.315	9.552	100	0.278	1.307	1.585	0.298	2.936	3.234
5	-10	100	0.086	1.862	1.948	100	0.087	1.863	1.949	100	0.087	1.863	1.949	100	7.357	24.080	31.437	22.995	0.878	2.098	2.976	0.070	3.181	3.251
	0	53.262	0.085	1.861	1.946	53.262	0.085	1.862	1.947	53.262	0.085	1.862	1.947	12.247	3.425	20.353	23.778	18.640	0.897	2.089	2.986	0.031	1.957	1.988
	10	100	0.085	1.862	1.947	100	0.086	1.862	1.948	100	0.086	1.862	1.948	9.928	0.836	9.263	10.099	65.707	0.917	2.085	3.001	0.062	3.224	3.285
6	-10	28.368	0.053	2.559	2.613	28.368	0.045	2.562	2.607	28.368	0.045	2.563	2.608	100	3.003	9.191	12.194	15.109	0.275	2.004	2.279	3.501	5.615	9.117
	0	3.474	0.052	0.456	0.508	3.474	0.051	0.452	0.503	1.850	0.001	0.073	0.074	4.286	0.034	0.340	0.374	1.5	0.216	0.917	1.133	96.237	2.672	98.909
	10	28.368	0.042	2.317	2.359	28.368	0.038	2.317	2.355	28.368	0.037	2.317	2.354	12.247	0.313	8.190	8.503	15.109	0.217	1.881	2.098	4.559	5.019	9.578
											Design 2													
1	-10	6.523	0.450	2.050	2.500	6.523	0.508	2.140	2.647	6.523	1.311	3.115	4.426	2.283	0.317	13.885	14.202	2.283	1.262	3.142	4.404	0.002	0.868	0.869
	0	12.247	2.569	4.920	7.489	12.247	2.156	5.046	7.202	12.247	1.631	5.238	6.870	65.707	0.028	10.381	10.409	9.928	0.763	3.090	3.853	0.027	10.650	10.677
	10	6.523	0.152	1.100	1.252	6.523	0.141	1.070	1.211	6.523	0.075	0.991	1.066	8.047	0.003	0.678	0.681	2.816	0.040	0.781	0.821	0.002	0.287	0.289
2	-10	28.368	0.020	3.054	3.074	28.368	0.001	3.198	3.199	28.368	0.028	3.304	3.331	28.368	17.188	21.477	38.665	9.928	0.336	4.188	4.524	9.033	4.628	13.661
	0	65.707	0.002	2.539	2.541	65.707	0.039	2.629	2.668	65.707	0.098	2.727	2.825	100	14.321	15.326	29.647	43.174	2.051	2.816	4.867	0.002	3.026	3.028
	10	28.368	0.050	2.867	2.917	34.996	0.272	2.694	2.967	34.996	0.172	2.781	2.952	100	3.640	15.059	18.699	22.995	0.000	2.757	2.757	9.541	4.091	13.632
3	-10	12.247	0.084	4.174	4.259	12.247	0.195	4.448	4.642	12.247	0.428	4.785	5.212	12.247	15.379	33.983	49.362	5.288	1.950	5.148	7.099	1.211	3.694	4.905
	0	65.707	0.008	2.653	2.660	81.06	0.044	2.752	2.796	65.707	0.112	2.827	2.939	100	9.828	13.330	23.158	43.174	1.372	2.790	4.162	0.003	4.172	4.175
	10	12.247	0.067	4.239	4.306	12.247	0.030	4.314	4.344	15.109	0.322	3.898	4.220	18.64	0.369	7.561	7.930	6.523	0.014	3.396	3.410	1.373	3.732	5.104
4	-10	81.060	0.004	1.661	1.665	81.06	0.076	1.670	1.746	100	0.137	1.707	1.843	100	14.109	3.821	17.930	34.996	1.013	1.381	2.394	0.008	2.785	2.793
	0	100	0.004	1.665	1.669	100	0.076	1.674	1.749	100	0.136	1.709	1.846	9.928	8.612	6.260	14.872	100	0.995	1.413	2.408	0.043	1.762	1.805
	10	100	0.005	1.670	1.675	100	0.075	1.679	1.754	100	0.136	1.714	1.849	12.247	4.237	5.315	9.552	34.996	0.999	1.404	2.404	0.026	3.131	3.157
5	-10	100	0.001	1.705	1.706	100	0.004	1.786	1.790	100	0.047	1.891	1.938	100	7.357	24.080	31.437	34.996	0.977	2.204	3.180	0.003	2.838	2.841
	0	65.707	0.001	1.704	1.705	53.262	0.003	1.781	1.785	43.174	0.045	1.886	1.931	12.247	3.425	20.353	23.778	34.996	0.966	2.227	3.193	0.011	1.763	1.774
	10	100	0.001	1.706	1.707	81.06	0.004	1.783	1.786	81.06	0.047	1.887	1.934	9.928	0.836	9.263	10.099	34.996	1.006	2.202	3.208	0.000	2.994	2.994
6	-10	28.368	0.564	2.688	3.253	28.368	0.288	2.726	3.014	28.368	0.193	2.744	2.936	100	3.003	9.191	12.194	15.109	0.024	2.240	2.264	5.284	5.818	11.102
	0	3.474	0.058	0.475	0.532	3.474	0.056	0.468	0.525	1.5	0.000	0.062	0.062	4.286	0.034	0.340	0.374	1.5	0.027	0.227	0.255	105.037	2.844	107.881
	10	28.368	0.524	2.479	3.003	28.368	0.308	2.485	2.793	28.368	0.200	2.553	2.753	12.247	0.313	8.190	8.503	15.109	0.023	2.056	2.079	6.593	5.043	11.636

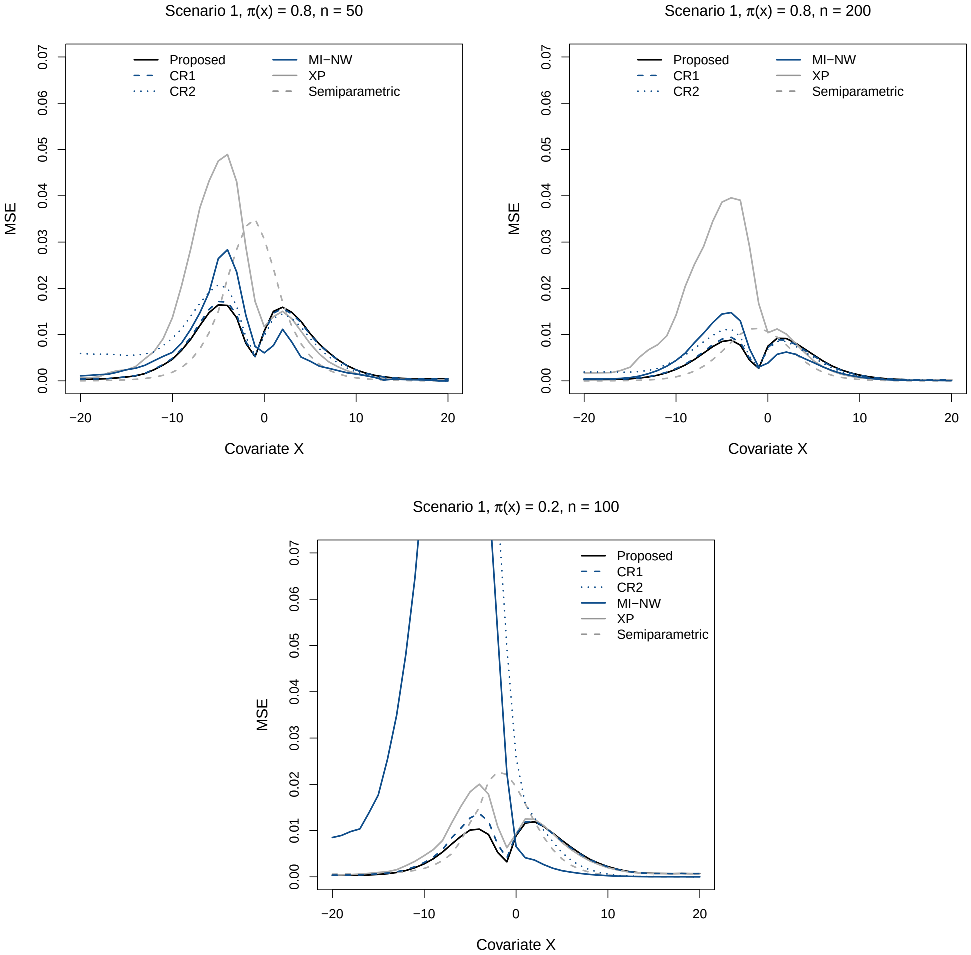

Figure 3.3: MSE of $1-\widehat{p}_h^c(x)$, $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$, $1-\widehat{p}_h^{\text{MI-NW}}(x)$, $1-\widehat{p}_h(x)$ (all computed with the optimal bandwidth), and $1-p(x;\widehat{\gamma})$ in Scenario 1 and under Design 2, for $\pi(x)=0.8, n=50, 200$ and $\pi(x)=0.2, n=100$.

Table 3.3: Squared bias (Bias^2), variance (Var) and MSE of $1-\widehat{p}_h^c(x)$, $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$, $1-\widehat{p}_h(x)$, $1-\widehat{p}_h^{\text{MI-NW}}(x)$ (all computed with the optimal bandwidth), and $1-p(x;\widehat{\boldsymbol{\gamma}})$ in Scenario 1 and under Design 2, for $\pi(x)=0.8$, and $n=50,100,200$.

		$1-\widehat{p}_h^c(x)$				$1-\widehat{p}_{1,h}(x)$				$1-\widehat{p}_{2,h}(x)$				$1-\widehat{p}_h(x)$				$1-\widehat{p}_h^{\text{MI-NW}}(x)$				$1-p(x;\widehat{\boldsymbol{\gamma}})$		
n	x	h_x	**Bias**^{2}	**Var**	**MSE**	h_x	**Bias**^{2}	**Var**	**MSE**	h_x	**Bias**^{2}	**Var**	**MSE**	h_x	**Bias**^{2}	**Var**	**MSE**	h_x	**Bias**^{2}	**Var**	**MSE**	**Bias**^{2}	**Var**	**MSE**
			$\times 10^3$	$\times 10^3$	$\times 10^3$		$\times 10^3$	$\times 10^3$	$\times 10^3$		$\times 10^3$	$\times 10^3$	$\times 10^3$		$\times 10^3$	$\times 10^3$	$\times 10^3$		$\times 10^3$	$\times 10^3$	$\times 10^3$	$\times 10^3$	$\times 10^3$	$\times 10^3$
50	-10	6.523	0.403	4.254	4.657	6.523	0.445	4.412	4.857	8.047	2.880	6.326	9.206	5.288	0.834	12.806	13.640	1.5	0.423	5.731	6.153	0.000	1.857	1.858
	0	18.64	4.169	6.571	10.740	22.995	4.748	5.665	10.413	22.995	3.837	5.851	9.688	65.707	0.103	11.604	11.707	12.247	0.020	6.036	6.056	0.369	30.307	30.676
	10	6.523	0.146	2.239	2.384	6.523	0.140	2.218	2.358	6.523	0.057	1.881	1.939	8.047	0.030	1.571	1.601	4.286	0.265	1.208	1.473	0.008	0.652	0.660
100	-10	6.523	0.450	2.050	2.500	6.523	0.508	2.140	2.647	6.523	1.311	3.115	4.426	2.283	0.317	13.885	14.202	2.283	1.262	3.142	4.404	0.002	0.868	0.869
	0	12.247	2.569	4.920	7.489	12.247	2.156	5.046	7.202	12.247	1.631	5.238	6.870	65.707	0.028	10.381	10.409	9.928	0.763	3.090	3.853	0.027	10.650	10.677
	10	6.523	0.152	1.100	1.252	6.523	0.141	1.070	1.211	6.523	0.075	0.991	1.066	8.047	0.003	0.678	0.681	2.816	0.040	0.781	0.821	0.002	0.287	0.289
200	-10	5.288	0.194	1.154	1.348	5.288	0.223	1.207	1.430	5.288	0.516	1.620	2.135	3.474	0.903	10.716	11.619	1.85	0.933	1.751	2.684	0.000	0.493	0.493
	0	9.928	1.569	3.005	4.574	9.928	1.241	3.058	4.298	9.928	1.003	3.078	4.081	100	0.010	5.969	5.979	8.047	0.322	1.660	1.982	0.018	4.762	4.780
	10	5.288	0.047	0.529	0.576	5.288	0.038	0.506	0.544	5.288	0.017	0.477	0.494	8.047	0.001	0.319	0.320	2.816	0.028	0.348	0.376	0.001	0.140	0.141

Table 3.4: Squared bias (Bias^2), variance (Var) and MSE of $1-\widehat{p}_h^c(x)$, $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$, $1-\widehat{p}_h(x)$, $1-\widehat{p}_h^{\text{MI-NW}}(x)$, all computed with the optimal bandwidth, and $1-p(x;\widehat{\boldsymbol{\gamma}})$ in Scenario 1 and under Design 2, for $\pi(x)=0.2, 0.8$, and $n=100$.

		$1-\widehat{p}_h^c(x)$				$1-\widehat{p}_{1,h}(x)$				$1-\widehat{p}_{2,h}(x)$				$1-\widehat{p}_h(x)$				$1-\widehat{p}_h^{\text{MI-NW}}(x)$				$1-p(x;\widehat{\boldsymbol{\gamma}})$		
$\pi(x)$	x	h_x	**Bias**^{2}	**Var**	**MSE**	h_x	**Bias**^{2}	**Var**	**MSE**	h_x	**Bias**^{2}	**Var**	**MSE**	h_x	**Bias**^{2}	**Var**	**MSE**	h_x	**Bias**^{2}	**Var**	**MSE**	**Bias**^{2}	**Var**	**MSE**
			$\times 10^3$	$\times 10^3$	$\times 10^3$		$\times 10^3$	$\times 10^3$	$\times 10^3$		$\times 10^3$	$\times 10^3$	$\times 10^3$		$\times 10^3$	$\times 10^3$	$\times 10^3$		$\times 10^3$	$\times 10^3$	$\times 10^3$	$\times 10^3$	$\times 10^3$	$\times 10^3$
0.2	-10	6.523	0.410	2.403	2.813	6.523	0.461	2.632	3.093	9.928	169.346	60.527	229.873	5.288	0.309	4.279	4.588	2.816	69.363	17.956	87.318	0.002	1.827	1.829
	0	15.109	3.639	5.259	8.899	18.64	3.837	5.308	9.146	100	10.318	15.544	25.862	22.995	2.903	6.399	9.301	100	2.327	4.234	6.560	1.027	18.455	19.481
	10	8.047	0.530	1.686	2.216	8.047	0.496	1.663	2.160	4.286	0.136	0.451	0.587	8.047	0.359	1.619	1.977	4.286	0.010	0.238	0.248	0.000	0.390	0.391
0.8	-10	6.523	0.450	2.050	2.500	6.523	0.508	2.140	2.647	6.523	1.311	3.115	4.426	2.283	0.317	13.885	14.202	2.283	1.262	3.142	4.404	0.002	0.868	0.869
	0	12.247	2.569	4.920	7.489	12.247	2.156	5.046	7.202	12.247	1.631	5.238	6.870	65.707	0.028	10.381	10.409	9.928	0.763	3.090	3.853	0.027	10.650	10.677
	10	6.523	0.152	1.100	1.252	6.523	0.141	1.070	1.211	6.523	0.075	0.991	1.066	8.047	0.003	0.678	0.681	2.816	0.040	0.781	0.821	0.002	0.287	0.289

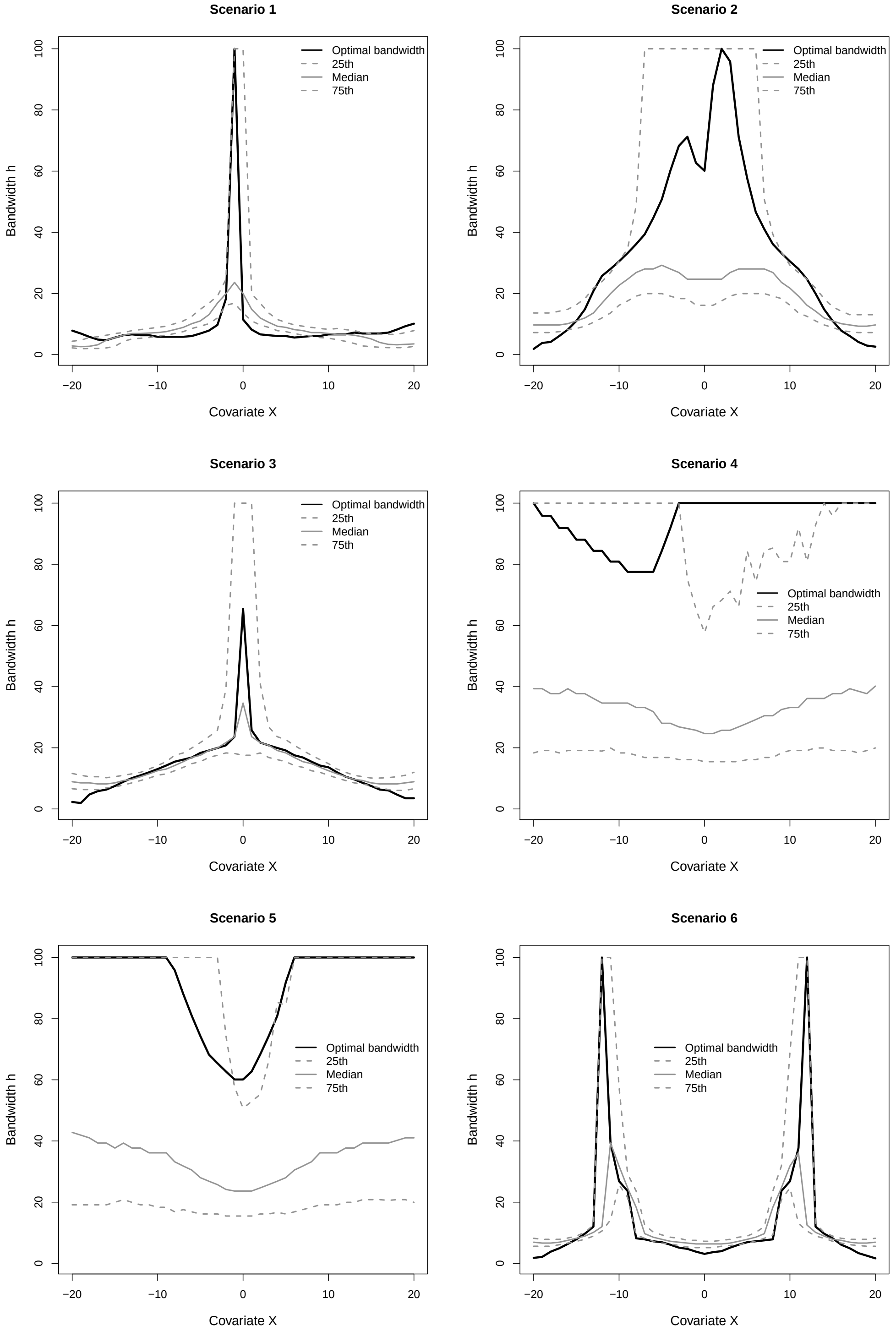

Figure 3.4: Median, first and third quartile of the bootstrap bandwidths for $1-\widehat{p}_h^c(x)$ in Scenarios 1−6 and under Design 2, for $\pi(x)=0.8$ and $n=100$. The optimal bandwidth (black solid line) is displayed for reference.

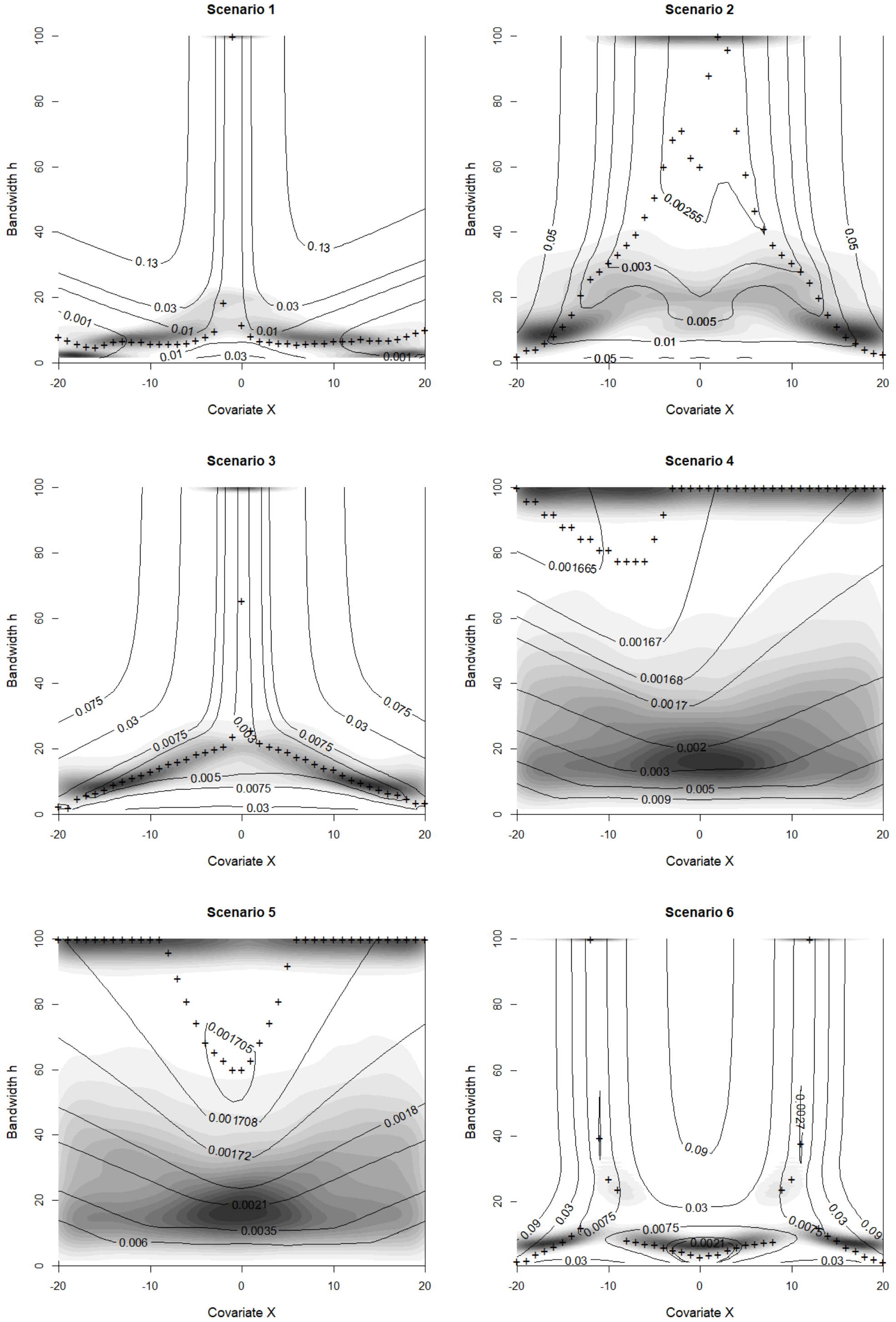

Figure 3.5: Contour plots of the MSE of $1-\widehat{p}^c_h(x)$ as a function of the bandwidth h and x in Scenarios 1–6 under Design 2 for $\pi(x) = 0.8$ and $n = 100$. For each x, the optimal bandwidth is marked with a cross. The density of the bootstrap bandwidths h^*_x is shown in gray shades (where a darker gray represents a higher density).

3.5 Real data analysis

To illustrate the practical performance of $1 - \widehat{p}_h^c(x)$ in (3.1) and $1 - \widehat{p}_n^c$ in (3.3), these estimators were applied to breast cancer and COVID-19 datasets described in Sections 1.6.

3.5.1 Breast cancer data

When analyzing the survival of the breast cancer patients, it is of great interest to study the clinical effect of well-established clinicopathologic prognostic factors (Duffy et al., 2017; Colomer et al., 2018) of the patients. The aim of the analysis presented here was to estimate the probability of not dying from breast cancer (probability of cure when the event of interest is death from breast cancer) depending on cancer stage, number of positive lymph nodes, menopausal status, margin status and age at diagnosis. In this dataset, only 42 (4.7%) patients died from cancer within the follow-up period. The observed times until death from breast cancer for the remaining patients were right-censored. In this censored group, 20 patients (2.2%) were cancer free for more than 10 years, suggesting they might be cured from the event "death because of cancer". This results in a very high missingness rate, 93.1%, for the cure indicator, ν. Maller and Zhou (1992) test (p-value < 0.001) provides evidence supporting condition (3.5), clearly set forth the use of $1 - \widehat{p}_h^c(x)$ and $1 - \widehat{p}_n^c$.
The probability of not dying from breast cancer $1 - p$ for different groups of patients according to the aforementioned covariates was estimated using:

(a) the estimator in (3.3) with its $\widehat{se}_B(1 - \widehat{p}_n^c)$;

(b) the empirical estimator $1 - \widehat{p} = \sum_{i=1}^n \xi_i \nu_i \left(\sum_{i=1}^n \xi_i\right)^{-1}$ with its $\widehat{se}(1 - \widehat{p})$, which discards the patients with unknown cure status;

(c) the unconditional competing risks estimators $1 - \widehat{p}_{1,n}$ with its $\widehat{se}_B(1 - \widehat{p}_{1,n})$, $1 - \widehat{p}_{2,n}$ with its $\widehat{se}_B(1 - \widehat{p}_{2,n})$;

(d) the MI estimator in (3.18) with its $\widehat{se}_B(1 - \widehat{p}_n^{\mathrm{MI}})$ and $M = 20$;

(e) the unconditional XP estimator $1 - \widehat{p}_n$ in (3.4) with its $\widehat{se}(1 - \widehat{p}_n)$.

The results are given in Table 3.5.

For consistency, standard errors $\widehat{se}_B(1-\widehat{p}_n^c)$, $\widehat{se}_B(1-\widehat{p}_{1,n})$, $\widehat{se}_B(1-\widehat{p}_{2,n})$, and $\widehat{se}_B(1-\widehat{p}_n^{\text{MI}})$ were computed using the bootstrap resampling procedure in Section 2.3.3. To the best of our knowledge, there is no any specifically tailored bandwidth selector for $1-\widehat{p}_n^{\text{MI}}$. Thus, in this analysis the pilot bandwidths g_1 and g_2 for Steps 1 and 2 were selected using the cross-validation selector of Bowman et al. (1998), available in an `R` package `kerdiest` (Quintela-del Río and Estévez-Pérez, 2012). The standard error $\widehat{se}(1-\widehat{p}_n)$ was computed with Greenwood's formula using the `R` package `survival`.

The empirical estimator $1-\widehat{p}$ seems to underestimate the true $1-p$. Note that all the patients with unknown cure status are excluded, so the estimate is computed with a considerably reduced sample size. If the excluded patients are not MCAR, the reduced sample might not be representative. The unconditional MI-NW estimator uses the complete sample as it takes into consideration the patients with unknown cure status, but it still appears to be performing poorly because 93.1% of patients have missing cure status. The unconditional XP estimator, $1-\widehat{p}_n$, does consider the censored observations, however, it dismisses the cure status information so it still underestimates the true cure probabilities. The estimators $1-\widehat{p}_n^c$, $1-\widehat{p}_{1,n}$ and $1-\widehat{p}_{2,n}$ make use of the available information of the cure status giving a reasonably accurate estimates.

The estimated probability of not dying from breast cancer as a function of a continuous covariate like age, is given in Figure 3.6. The estimator $1-\widehat{p}_h^c(x)$ is compared with the competing risks estimators $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$, and the XP estimator $1-\widehat{p}_h(x)$, all computed with the bootstrap bandwidth selector discussed in Section 3.2.3 using $B = 1000$ resamples. It is also compared with the semiparametric estimator $1-p(x;\widehat{\boldsymbol{\gamma}})$ and the MI-NW estimator $1-\widehat{p}_h^{\text{MI-NW}}(x)$ computed with $M = 20$. Note that the bandwidth h for the MI-NW estimator was chosen via an improved cross-validation bandwidth selector for the NW estimator (Hurvich et al., 1998), using the `R` package `np` (Tristen and Jeffrey, 2008).

Figure 3.6 also shows the 95% CI of the cure probabilities derived by the estimator $1-\widehat{p}_h^c(x)$. To estimate these CI, the bootstrap procedure similar to Section 2.5 was used to estimate the standard error of $1-\widehat{p}_h^c(x)$, $\widehat{se}_B\left(1-\widehat{p}_h^c(x)\right)$ with $B = 1000$ bootstrap resamples. The $100(1-\alpha)$% CI of $1-\widehat{p}_h^c(x)$ is estimated as

$$1-\widehat{p}_h^c(x) \mp z_{1-\frac{\alpha}{2}}\widehat{se}_B\left(1-\widehat{p}_h^c(x)\right),$$

where z_β is the βth quantile of the standard normal.

Although $1-p(x;\widehat{\boldsymbol{\gamma}})$ shows that the probability of not dying from breast can-

Table 3.5: Demographic characteristics of breast cancer patients, and the numbers of dead patients (Death), patients known to be cured (Cured) and patients with unknown cure status (Unknown). Also given are the estimated probability of being cured from breast cancer $(1-p)$ estimated using $1-\widehat{p}_n^c$, $1-\widehat{p}$, $1-\widehat{p}_{1,n}$, $1-\widehat{p}_{2,n}$, $1-\widehat{p}_n^{\mathrm{MI}}$, and $1-\widehat{p}_n$ with their respective estimated standard errors (se).

Characteristic	**Count** (%)	**Uncured**	**Censored**		**Estimated probability of not dying from cancer**					
		Dead	**Cured**[‡]	**Unknown**	$1-\widehat{p}_n^c$ $(\widehat{se}_B)$	$1-\widehat{p}$ $(\widehat{se})$	$1-\widehat{p}_{1,n}$ $(\widehat{se}_B)$	$1-\widehat{p}_{2,n}$ $(\widehat{se}_B)$	$1-\widehat{p}_n^{\mathrm{MI}}$ $(\widehat{se}_B)$	$1-\widehat{p}_n$ $(\widehat{se})$
Age										
<55 years	371 (41.3)	21	11	339	0.579 (0.084)	0.344 (0.084)	0.579 (0.084)	0.579 (0.099)	0.363 (0.093)	0.561 (0.097)
≥ 55 years	527 (58.7)	21	9	497	0.766 (0.089)	0.300 (0.084)	0.766 (0.089)	0.766 (0.089)	0.310 (0.095)	0.766 (0.081)
Stages[†]										
I	164 (19.8)	5	4	155	0.812 (0.099)	0.444 (0.166)	0.812 (0.099)	0.812 (0.143)	0.372 (0.191)	0.812 (0.092)
II	514 (62.0)	20	10	484	0.630 (0.095)	0.333 (0.086)	0.631 (0.095)	0.631 (0.103)	0.352 (0.098)	0.590 (0.118)
III	151 (18.2)	10	4	137	0.593 (0.148)	0.286 (0.121)	0.593 (0.148)	0.474 (0.170)	0.308 (0.136)	0.593 (0.139)
Menopausal [†] **status**[†]										
Pre	185 (22.3)	10	0	175	0.248 (0.244)	0.000 (0.000)	0.248 (0.244)	0.000 (0.000)	0.000 (0.000)	0.248 (0.205)
Peri	63 (7.6)	7	11	45	0.725 (0.094)	0.611 (0.115)	0.725 (0.094)	0.725 (0.105)	0.596 (0.121)	0.707 (0.105)
Post	581 (70.1)	17	9	555	0.803 (0.091)	0.346 (0.093)	0.803 (0.091)	0.723 (0.109)	0.397 (0.107)	0.803 (0.083)
No. of positive lymph nodes[†]										
0	392 (50.5)	12	11	369	0.754 (0.085)	0.478 (0.104)	0.754 (0.085)	0.754 (0.085)	0.484 (0.115)	0.754 (0.087)
1–3	299 (38.5)	16	6	277	0.639 (0.118)	0.273 (0.095)	0.639 (0.118)	0.548 (0.139)	0.337 (0.111)	0.639 (0.114)
>3	86 (11.1)	11	3	72	0.468 (0.152)	0.214 (0.110)	0.468 (0.152)	0.351 (0.150)	0.270 (0.124)	0.439 (0.157)
Margin status[†]										
Negative	761 (92.5)	21	14	726	0.744 (0.075)	0.400 (0.083)	0.744 (0.075)	0.744 (0.090)	0.406 (0.092)	0.730 (0.083)
Positive	62 (7.5)	4	3	55	0.771 (0.131)	0.429 (0.187)	0.771 (0.131)	0.771 (0.184)	0.431 (0.216)	0.771 (0.128)

[†] Missing observations present

[‡] “Cured” from the event “death from breast cancer”

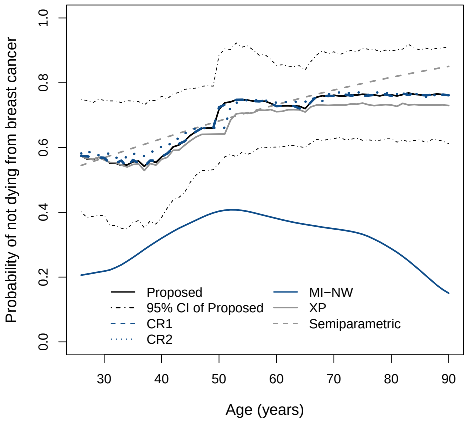

Figure 3.6: Estimation of the probability of not dying from breast cancer patients by using $1-\widehat{p}^{c}_{h}(x)$ and its 95% CI, $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$, $1-\widehat{p}_{h}(x)$ (all computed with the bootstrap bandwidth), $1-\widehat{p}^{\text{MI-NW}}_{h}(x)$ (computed using the cross-validation bandwidth), and $1-p(x;\widehat{\boldsymbol{\gamma}})$.

cer increases with age, the curves from the other estimators suggest that the logistic model assumed in the semiparametric estimator might not be appropriate. Specifically, they indicate an increment of that probability only for younger to middle age patients. The estimators $1-\widehat{p}^{c}_{h}(x)$, $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$, and $1-\widehat{p}_{h}(x)$ suggest no effect of the age on the probability for elderly patients, while $1-\widehat{p}^{\text{MI-NW}}_{h}(x)$ implies that the probability decreases with the age in older patients.

Observe that the probability of not dying from breast cancer given by $1-\widehat{p}_{h}(x)$, an estimator that disregards the available information about the cure status, is equal or lower than the probability estimated with $1-\widehat{p}^{c}_{h}(x)$. This means that the probability of not dying from breast cancer is likely to be underestimated by $1-\widehat{p}_{h}(x)$. Nonetheless, the differences between $1-\widehat{p}^{c}_{h}(x)$ and $1-\widehat{p}_{h}(x)$ are subtle, as the proportion of the identified known cures is small. When the last observation is an event or known to be cured, the estimator $1-\widehat{p}^{c}_{h}(x)$ produces the same estimate as the competing risks estimators $1-\widehat{p}_{1,h}(x)$ and $1-\widehat{p}_{2,h}(x)$.

Finally, regarding the MI-NW estimator, it shows a similar trend as $1-\widehat{p}^{c}_{h}(x)$ and $1-\widehat{p}_{h}(x)$, although the estimated probabilities are substantially smaller. As pointed out before, the performance of $1-\widehat{p}^{\text{MI-NW}}_{h}(x)$ worsens significantly because of the extremely high proportion of patients with missing cure status.

3.5.2 COVID-19 data

Since the COVID-19 pandemic started in the beginning of 2020, countries around the world are experiencing a large number of incident cases, with many patients requiring hospitalization wards. Although most infected people presented with mild disease, there were many severe cases that required long stays in ICU, overwhelming the healthcare systems with critical consequences on the disease mortality. An accurate knowledge of the duration of hospitalization, and the prediction of the probability that a hospitalized inpatient would require a bed in ICU, are key for understanding the hospital demand for beds and crucial for decision-making and suitable planning.

As mentioned in Section 1.6, the second dataset contains the $10,454$ confirmed COVID-19 cases reported by the Galician Healthcare Service (2021) between March 6 and May 7, 2020. The time of interest is the length of stay in hospital ward until admission to ICU, and the aim of this analysis was to estimate the probability of admission to ICU from hospital ward given age and sex as covariates of interest, see Table 3.6. Of the $2,484$ hospitalized cases, 104 (4.2%) patients were excluded from analysis because they were admitted and discharged on the same date or they were admitted directly to the ICU, resulting in a length of stay in hospital ward of 0 days. For the remaining $2,380$ hospitalized patients for at least one day, $1,063$ (44.7%) were 75 years of age or above and $1,262$ (53.0%) were males. A total of $1,638$ (68.8%) patients were discharged alive before entering ICU, and 328 (13.8%) had died before entering ICU. None of them will require admission to ICU eventually, so all of them can be considered as “cured” from the event of interest, that is, admission to ICU. Note that “cure” means being free of experiencing admission to ICU, not cured in medical terms.

A total of 197 patients of the $2,380$ inpatients in hospital ward required admission to ICU, which gives an empirical estimated probability of admission to ICU of $\hat{p} = 197/2380 = 0.083$. However, the true number of patients requiring ICU is expected to be larger than 197, as some of the 217 (9.1%) inpatients still in hospital bed at the end of the study might eventually need admission to the ICU. This shows that $\hat{p} = 0.083$ might underestimate the probability of admission to ICU, motivating the use of alternative estimators than can handle censoring such as the proposed estimator. It is assumed that condition (3.5) applies, as the result of the test of Maller and Zhou (Maller and Zhou, 1992) suggests (p-value < 0.001).

Table 3.6: Demographic characteristics of COVID-19 patients, and the numbers of dead patients (Death), patients known to be cured (Cured) and patients with unknown cure status (Unknown). Also given are the estimated probability of admission to ICU (p), when the probability of not requiring ICU $(1-p)$ is estimated using $1-\widehat{p}_n^{\mathrm{c}}$, $1-\widehat{p}$, $1-\widehat{p}_{1,n}$, $1-\widehat{p}_{2,n}$, $1-\widehat{p}_n^{\mathrm{MI}}$ and $1-\widehat{p}_n$ with their respective estimated standard errors (se).

		Uncured	**Censored**			**Estimated probability of requiring ICU**					
Variables	**Count** (%)	**ICU admission**	**Dead**‡	**Discharged**‡	**Unknown**	$\widehat{p}_n^{\mathrm{c}}$ $(\widehat{se}_B)$	$\widehat{p}$ $(\widehat{se})$	$\widehat{p}_{1,n}$ $(\widehat{se}_B)$	$\widehat{p}_{2,n}$ $(\widehat{se}_B)$	$\widehat{p}_n^{\mathrm{MI}}$ $(\widehat{se}_B)$	$\widehat{p}_n$ $(\widehat{se})$
Age											
0 – 24 years	22 (0.9)	0	0	17	5	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.156 (0.087)	0.000 (0.000)	0.000 (0.000)
25 – 54 years	359 (15.1)	25	5	316	13	0.071 (0.013)	0.072 (0.014)	0.000 (0.014)	0.071 (0.014)	0.077 (0.014)	0.172 (0.056)
55 – 64 years	354 (14.9)	41	24	271	18	0.114 (0.017)	0.122 (0.018)	0.117 (0.017)	0.123 (0.018)	0.121 (0.018)	0.138 (0.022)
65 – 74 years	582 (24.5)	96	35	411	40	0.165 (0.016)	0.177 (0.016)	0.169 (0.016)	0.180 (0.018)	0.176 (0.017)	0.318 (0.076)
75 – 84 years	571 (24.0)	34	101	377	59	0.059 (0.010)	0.066 (0.011)	0.061 (0.010)	0.061 (0.011)	0.067 (0.011)	0.083 (0.016)
85 years and over	492 (20.7)	1	163	246	82	0.002 (0.002)	0.002 (0.002)	0.002 (0.002)	0.027 (0.012)	0.003 (0.003)	0.004 (0.004)
Sex											
Female	1118 (47.0)	55	136	822	105	0.049 (0.006)	0.054 (0.007)	0.051 (0.007)	0.051 (0.007)	0.053 (0.007)	0.091 (0.022)
Male	1262 (53.0)	142	192	816	112	0.114 (0.009)	0.124 (0.010)	0.115 (0.009)	0.131 (0.012)	0.124 (0.010)	0.175 (0.034)

‡ "cured" from the event "admission to ICU"

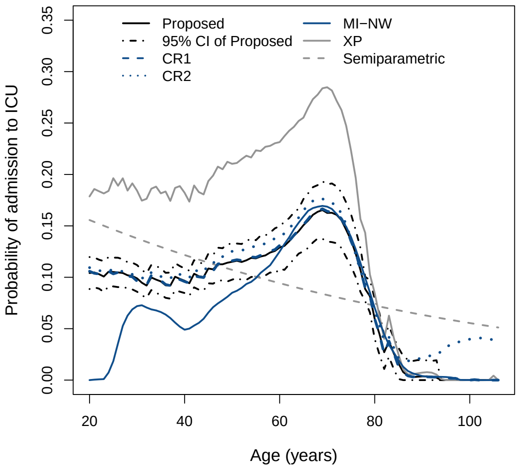

Figure 3.7: Estimation of the probability of admission to ICU for hospitalized COVID-19 patients estimated using $1 - \widehat{p}^{\text{c}}_h(x)$ and its 95% CI, $1 - \widehat{p}_{1,h}(x)$, $1 - \widehat{p}_{2,h}(x)$, $1 - \widehat{p}_h(x)$ (all computed with the bootstrap bandwidth), $1 - \widehat{p}^{\text{MI-NW}}_h(x)$ (computed using the cross-validation bandwidth), and $1 - p(x; \widehat{\boldsymbol{\gamma}})$.

Table 3.6 shows the estimated probabilities p of requiring ICU, given by the proposed estimator $\widehat{p}^{\text{c}}_n$, the unconditional competing risks estimators $\widehat{p}_{1,n}$, $\widehat{p}_{2,n}$, the empirical estimator $\widehat{p}$, the MI estimator $\widehat{p}^{\text{MI}}_n$ with $M = 20$, and the unconditional XP estimator $\widehat{p}_n$. It should be noted that only 9.1% patients are still in hospital bed at the end of the study, for whom eventual admission to ICU is unknown (missing cure status). Therefore, the proportion of individuals with the observed cure status is high.

In this situation, the estimators $\widehat{p}^{\text{c}}_n$, $\widehat{p}_{1,n}$ and $\widehat{p}^{\text{MI}}_n$ are expected to perform nicely, and the results by $\widehat{p}$ are likely to improve as the biased performance towards insufficient cure status information fades away. On the other hand, the estimator $\widehat{p}_{2,n}$ tends to overestimate low probability of requiring ICU. In addition, XP estimator is expected to perform poorly since it dismisses the significant information given by the observed cured individuals.

As it can be seen in Table 3.6, the estimated probabilities of admission to ICU given by the empirical estimator, the MI estimator and the proposed estimator are very similar. This suggests that it is possible that the missing data mechanism from this dataset is close to strongly ignorable missing at random. On the other hand, the estimated probabilities given by the XP estimator $\widehat{p}_n$ seem to be too high.

Figure 3.7 shows the estimated probability of requiring admission to ICU depending on the age, obtained using the estimator $1-\widehat{p}_h^c(x)$, the competing risks estimators $1-\widehat{p}_{1,h}(x)$, $1-\widehat{p}_{2,h}(x)$, the XP estimator $1-\widehat{p}_h(x)$, all computed using the bootstrap bandwidth selector as in the breast cancer example, the semiparametric estimator $1-p(x;\widehat{\gamma})$, and the MI-NW estimator $1-\widehat{p}_h^{\text{MI-NW}}(x)$ computed using the same bandwidths selectors as in the breast cancer example. Although the semiparametric estimator suggests a uniformly decreasing effect of the age on the probability of admission to the ICU, the other three estimators indicate that the logistic assumption for the cure probability might not be acceptable, as the curve patterns are characterized by a constant to a slightly increasing probability of admission to the ICU for younger patients (below 55 years), a sharp increase of the probability for middle age patients (from 55 to 69 years) and a decrease for elderly patients (70 years or older). For the aforementioned reasons, the XP estimator seems to overestimate the probability of ICU admission, now as a function of the age. Regarding the MI-NW estimator, the pattern of the estimated probability is consistent with that of the proposed estimator. However, it seems to underestimate the probability of admission to ICU for young-to-middle age patients. This is due to the low percentage of observed admissions to ICU in patients of those ages, resulting in an estimation with a high percentage of missing response.

3.6 Discussion

A novel nonparametric estimator of the conditional probability of cure is proposed for the MCM when some censored individuals can be observed to be cured from the event. It reduces to well-known estimators in the literature for the unconditional setting, when there is a cure threshold, if there is no observed cured individuals, and when there is no censoring. In contrast to regression based estimators, the proposed estimator is based on the MCM. It uses the available information of the observed times, and therefore can lead to substantial gain in efficiency.

When compared with the XP estimator, also based on the MCM but disregarding the information given by the cure status, it has been demonstrated that the estimator of the probability of cure proposed in Section 3.2 has always smaller asymptotic variance. The advantage in terms of bias is not guaranteed, as it depends on the conditional probability of knowing the cure status information and the censoring distribution. On the other hand, simulations and the analysis of two real data examples show that our estimator yields significant improvement compared to XP estimator, for which the cure status information is ignored.

In Section 3.3.1, the cure rate can alternatively be estimated using a competing risks model, with the event as the type of failure of interest and the cure as a competing risk. The basic assumption in this competing risks model is that failure, cure, and censoring variables are independent conditional on the covariate, and that either failure or cure will occur with probability one although they may not be observed (Betensky and Schoenfeld, 2001). The main disadvantage of this approach is that if the last observation is not an event nor an observed cured individual, then the estimator of the cure rate is not unique, and only upper and lower bounds are provided.

The multiply imputed NW estimator, introduced in Section 3.3.2, performs well when the proportion of observed events is high, that is, for low percentages of missing events, but it performs poorly when there is heavy missingness. In this setup, high levels of missingness are linked to low values of $\pi(x) = P(C = \infty \mid X = x)$, which results in a clear violation of the siMAR assumption, and consequently biased estimates. Besides, the MI-MW estimator is computationally quite expensive, particularly when the sample size is large, and it requires the selection of three different bandwidths.

It should be noted that the semiparametric estimator by Bernhardt (2016) is somewhat affected when the logistic assumption is violated and it might be challenging to obtain stable estimates for the model parameters if the sample size is small.

Finally, the empirical estimator of the unconditional cure probability, which disregards the observations with unknown cure status, clearly underestimates the true probability and it cannot handle continuous covariates.

As discussed in Section 3.3.2, the probability of cure can be estimated as a regression function with the cure indicator as the response. The proposed MI-NW estimator imputes the missing values of the cure status. We are aware of the existence of other ways of dealing with missingness in the response. To name one of them, the problem can also be addressed using inverse probability weighting method (Wang et al., 2010; Seaman et al., 2012; Seaman and White, 2013). However, it has not been considered in this chapter for the estimation of the cure rate because the performance is expected to be similar to that of the MI-NW estimator, and the efficiency of this method is expected to depend largely on the level of missingness in the data.

In summary, the proposed estimator in Section 3.2 performs well when there are individuals known to be "cured" from the considered event, and is efficient under high proportion of missingness in the cure status. Moreover, it does not assume any parametric assumption for the cure probability. Also, it can be applied when the cure identification does not rely necessarily on the observed time-to-event being larger than a cure threshold, when there is not any individual known to be cured or without censoring.

Chapter 4

Nonparametric latency estimator

4.1 Introduction

While in Chapter 3 the interest was on estimating the probability of cure or its complement (the probability of the event), the focus of this chapter is on estimating the latency function. The latency function can be estimated using either a (semi)parametric approach (Maller and Zhou, 1992; Patilea and Van Keilegom, 2020; Amico and Van Keilegom, 2018, among others) or a completely nonparametric method (López-Cheda et al., 2017b). Peng and Yu (2021) provide a comprehensive discussion on the different procedures for estimating the latency function in the standard MCM.

Few authors have developed methods to estimate the latency function when the cure status is available in the MCM. Wu et al. (2014) modeled the latency part using Cox proportional hazards model, whereas Bernhardt (2016) considered the AFT model. Completely nonparametric estimation of the latency function, without covariates, has been addressed by Laska and Meisner (1992).

The aim of this chapter is to develop a fully nonparametric estimator of the conditional latency function in the MCM when the cure status is partially available. The proposed estimator extends the nonparametric latency estimator of López-Cheda et al. (2017b). This chapter proceeds as follows. In Section 4.2, a nonparametric estimator of the conditional latency function is proposed and the asymptotic properties of the estimator are studied. Also, a bootstrap bandwidth selector is proposed. In Section 4.3, the results of a simulation study carried out to evaluate the finite sample performance of the estimator are presented. An application to the COVID-19 data is delineated in Section 4.4. A discussion is given in Section 4.5.

4.2 Proposed estimator of the latency function

From the definition of the survival function in MCM, given in (1.4), the latency function can be written in terms of the survival function and the cure probability as follows:

$$S_0\left(t \mid x\right) = \frac{S\left(t \mid x\right) - (1 - p(x))}{p(x)}. \tag{4.1}$$

Following similar ideas as in López-Cheda et al. (2017b), equation (4.1) is used to motivate a nonparametric estimator of the latency function, by replacing $S\left(t \mid x\right)$ and $1 - p(x)$ with the estimators $\widehat{S}_{h_2}^c\left(t \mid x\right)$ in (2.16) and $1 - \widehat{p}_{h_1}^c\left(x\right)$ in (3.1), respectively, where the bandwidths h_1 and h_2 are allowed to be distinct. To estimate $S_0\left(t \mid x\right)$, the estimator below is proposed:

$$\widehat{S}_{0,h_1,h_2}^c\left(t \mid x\right) = \begin{cases} \dfrac{\widehat{S}_{h_2}^c\left(t \mid x\right) - (1 - \widehat{p}_{h_1}^c(x))}{\widehat{p}_{h_1}^c(x)} & \text{if } \ 0 \leq t \leq T_{(n)}^1 \ \text{ and} \\ & \quad \widehat{S}_{h_2}^c\left(t \mid x\right) > 1 - \widehat{p}_{h_1}^c(x) \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

Clearly, the condition $\widehat{S}_{h_2}^c\left(t \mid x\right) > 1 - \widehat{p}_{h_1}^c(x)$ in (4.2) is added to ensure the non-negativity of the estimator.

Note that if $h_1 = h_2 = h$ then the proposed estimator in (4.2) reduces to the following estimator:

$$\widehat{S}_{0,h}^c(t \mid x) = \frac{\widehat{S}_h^c\left(t \mid x\right) - (1 - \widehat{p}_h^c(x))}{\widehat{p}_h^c(x)}. \tag{4.3}$$

Although the estimator in (4.3) has the advantage of providing legitimate estimates of a survival function with a more straightforward definition than that of the estimator in (4.2), it might not be flexible enough when the optimal bandwidths for $\widehat{S}_h^c\left(t \mid x\right)$ and $1 - \widehat{p}_h^c(x)$ are quite different. Both estimators (4.2) and (4.3) extend the method in Laska and Meisner (1992) for randomly observed cured individuals.

In the unconditional case, the estimator in (4.2) becomes

$$\widehat{S}_{0,n}^c\left(t\right) = \frac{\widehat{S}_n^c\left(t\right) - (1 - \widehat{p}_n^c)}{\widehat{p}_n^c}, \tag{4.4}$$

where $\widehat{S}_n^c\left(t\right)$ is given in (2.18) and $1 - \widehat{p}_n^c$ in (3.3).

4.2.1 Asymptotic results

In this section, the asymptotic properties of the estimator $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ are established. From them, the asymptotic properties of the estimator with a single bandwidth in (4.3) are immediately derived by just considering $h_1 = h_2 = h$. The next theorem gives an asymptotic representation for $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$. Its proof is provided in Appendix C.

Theorem 4.1 (Asymptotic representation). Suppose that Assumptions 1−9 hold, then, for $x \in I$ and $t \in [a,b]$ such that $\widehat{S}^c_{h_2}(t \mid x) > 1 - \widehat{p}^c_{h_1}(x)$, an iid representation for $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ is

$$\widehat{S}^c_{0,h_1,h_2}(t \mid x) - S_0(t \mid x) = \sum_{i=1}^n \eta_{h_1,h_2}(T_i, \delta_i, \xi_i, \nu_i, t, x) + R_n(t,x), \tag{4.5}$$

where

$$\begin{aligned}\eta_{h_1,h_2}(T_i, \delta_i, \xi_i, \nu_i, t, x) = &- \frac{S(t \mid x)}{p(x)} \widetilde{B}_{h_2 i}(x) \zeta(T_i, \delta_i, \xi_i, \nu_i, t, x) \\ &- \frac{(1-p(x))(1-S(t \mid x))}{p^2(x)} \widetilde{B}_{h_1 i}(x) \zeta(T_i, \delta_i, \xi_i, \nu_i, \tau_0, x)\end{aligned} \tag{4.6}$$

with $\zeta(T_i, \delta_i, \xi_i, \nu_i, t, x)$ in (2.21),

$$\widetilde{B}_{h_j i}(x) = \frac{1}{m(x)} \frac{1}{n h_j} K\left(\frac{x - X_i}{h_j}\right), \text{ for } j = 1, 2,$$

and $R_n(t,x)$ can be shown to satisfy

$$\sup_{a \le t \le b, x \in I} | R_n(t,x) | = O\left((nh)^{-3/4} (\log n)^{3/4}\right) \text{ a.s.}$$

In the next proposition, the asymptotic bias and variance of the dominant term in the iid representation of $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ are studied.

Proposition 4.1 (**Asymptotic expression of the bias and variance**) Suppose that Assumptions 1−9 are satisfied, then, the asymptotic bias and variance of the dominant term of $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ are, respectively,

$$\mu^c_{h_1,h_2}(t,x) = h_1^2 B_{c,1}(t,x) + h_2^2 B_{c,2}(t,x) + O\left(h_1^4\right) + O\left(h_2^4\right), \tag{4.7}$$

and

$$\begin{aligned}\sigma^2_{c,h_1,h_2}(t,x) = &\frac{1}{nh_1} s^2_{c,1}(x) + \frac{1}{nh_2}\left(s^2_{c,2}(t,x) + 2 s^2_{c,3}(t,x)\right) \\ &+ O(n^{-1} h_2) + O((nh_2)^{-1} h_1).\end{aligned} \tag{4.8}$$

The dominant terms in the bias given in (4.7) are

$$B_{c,1}(t,x) = \frac{1 - S(t \mid x)}{2p^2(x) m(x)} d_K \left[2(1 - p(x))' m'(x) + (1 - p(x))'' m(x)\right]$$

$$-\frac{(1-p(x))(1-S(t\mid x))}{p^2(x)}d_K\int_0^{\tau_0}\frac{G'(v^-\mid x)}{1-G(v^-\mid x)}\frac{d}{ds}\left(\frac{S'(s\mid x)}{S(s\mid x)}\right)\bigg|_{s=v^-}dv \tag{4.9}$$

and

$$\begin{aligned}B_{c,2}(t,x)=&\frac{d_K}{2p(x)m(x)}\left(2S'\left(t^-\mid x\right)m'(x)+S''\left(t^-\mid x\right)m(x)\right)\\&-\frac{S(t\mid x)d_K}{p(x)}\int_0^t\frac{G'(v^-\mid x)}{1-G(v^-\mid x)}\frac{d}{ds}\left(\frac{S'(s\mid x)}{S(s\mid x)}\right)\bigg|_{s=v^-}dv,\end{aligned}\tag{4.10}$$

where $d_K=\int v^2K(v)dv$ and $S'(t\mid x), S''(t\mid x)$, $(1-p(x))'$ and $(1-p(x))''$ refer to the derivatives with respect to x.

The dominant terms in the variance given in (4.8) are

$$s^2_{c,1}(t,x)=\frac{(1-p(x))^2(1-S(t\mid x))^2}{p^4(x)m(x)}\int_0^{\tau_0}\frac{dH^1(v\mid x)}{(1-H(v\mid x)+H^{11}(v\mid x))^2}c_K \tag{4.11}$$

$$s^2_{c,2}(t,x)=\frac{S^2(t\mid x)}{p^2(x)m(x)}\int_0^t\frac{dH^1(v\mid x)}{(1-H(v\mid x)+H^{11}(v\mid x))^2}c_K \tag{4.12}$$

$$\begin{aligned}s^2_{c,3}(t,x)=&\frac{(1-p(x))(1-S(t\mid x))S(t\mid x)}{p^3(x)m(x)}\int_0^t\frac{dH^1(v\mid x)}{(1-H(v\mid x)+H^{11}(v\mid x))^2}\\&\times\int K(v)K(v\frac{h_1}{h_2})dv,\end{aligned}\tag{4.13}$$

where $c_K=\int K^2(v)dv$.

The proof of Proposition 4.1 is given in Appendix C.

Remark 4.1 Up to a factor depending on t and x, the terms $B_{c,1}(t,x)$ in (4.9) and $B_{c,2}(t,x)$ in (4.10) are the dominant terms of the bias of the estimators $\widehat{S}^c_{h_2}(t\mid x)$ and $1-\widehat{p}^c_{h_1}(x)$ given in (2.25) and (3.7), respectively. Similarly for the variance, the term $s^2_{c,1}(t,x)$ in (4.11) corresponds to the asymptotic variance of $1-\widehat{p}^c_{h_1}(x)$ in (3.10), and $s^2_{c,2}(t,x)$ in (4.12) corresponds to the asymptotic variance of $\widehat{S}^c_{h_2}(t\mid x)$ in (2.26). The last term $s^2_{c,3}(t,x)$ in (4.13) stands for the covariance of $1-\widehat{p}^c_{h_1}(x)$ and $\widehat{S}^c_{h_2}(t\mid x)$.

Theorem 4.2 (Asymptotic normality). Suppose that Assumptions 1−9 and 11 are satisfied, then, for $x\in I$ and $t\in[a,b]$ such that $\widehat{S}^c_{h_2}(t\mid x)>1-\widehat{p}^c_{h_1}(x)$, it follows that

(i) If $nh_1^5\to 0$ and $nh_2^5\to 0$, then

$$\sqrt{\frac{nh_1h_2}{h_1+h_2}}\left(\widehat{S}^c_{0,h_1,h_2}(t\mid x)-S_0(t\mid x)\right)\xrightarrow{d}N(0,s^2_c(t,x)),$$

where

$$s_c^2(t,x) = \begin{cases} s_{c,1}^2(t,x) & \text{if } \dfrac{h_1}{h_2} \to 0 \\[2ex] s_{c,2}^2(t,x) & \text{if } \dfrac{h_2}{h_1} \to 0 \\[2ex] \dfrac{C_2 s_{c,1}^2(t,x)}{C_1+C_2} + \dfrac{C_1\Big(s_{c,2}^2(t,x) + 2s_{c,3}^2(t,x)\Big)}{C_1+C_2} & \text{if } \dfrac{h_1}{h_2} \to \dfrac{C_1}{C_2} \end{cases} \tag{4.14}$$

with $s_{c,1}^2(t,x)$, $s_{c,2}^2(t,x)$ and $s_{c,3}^2(t,x)$ given in (4.11) – (4.13), and C_1 and C_2 are constants.

(ii) If $nh_1^5 \to 0$ and $nh_2^5 \to C_2^5 > 0$, then

$$\sqrt{\frac{nh_1h_2}{h_1+h_2}}\left(\widehat{S}^c_{0,h_1,h_2}(t \mid x) - S_0(t \mid x)\right) \xrightarrow{d} N(0, s_{c,1}^2(t,x)).$$

(iii) If $nh_1^5 \to C_1^5 > 0$ and $nh_2^5 \to 0$, then

$$\sqrt{\frac{nh_1h_2}{h_1+h_2}}\left(\widehat{S}^c_{0,h_1,h_2}(t \mid x) - S_0(t \mid x)\right) \xrightarrow{d} N(0, s_{c,2}^2(t,x)).$$

(iv) If $nh_1^5 \to C_1^5 > 0$ and $nh_2^5 \to C_2^5 > 0$, then

$$\sqrt{\frac{nh_1h_2}{h_1+h_2}}\left(\widehat{S}^c_{0,h_1,h_2}(t \mid x) - S_0(t \mid x)\right) \xrightarrow{d} N(B_c(t,x), s_c^2(t,x)),$$

where

$$B_c(t,x) = \sqrt{\frac{C_1C_2}{C_1+C_2}}\left(C_1^2 B_{c,1}(t,x) + C_2^2 B_{c,2}(t,x)\right),$$

with $B_{c,1}(t,x)$ and $B_{c,2}(t,x)$ defined in (4.9) and (4.10), and $s_c^2(t,x)$ is given in (4.14).

The proof of Theorem 4.2 is in Appendix C.

4.2.2 Effect of ignoring the cure status information

In this section, the effect of ignoring the cure status information is discussed by comparing the dominant terms of the bias and variance of the proposed estimator, $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$, with those of the LC estimator, $\widehat{S}_{0,h}(t \mid x)$ given in (1.6). The LC estimator is a particular case of the proposed estimators $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ and $\widehat{S}^c_{0,h}(t \mid x)$, only when a single bandwidth is used and any possibly available information about cure status is not considered in the estimation procedure.

Theorem 2 in López-Cheda et al. (2017b) presents asymptotic expressions for the bias and variance of the LC estimator, which are, respectively,

$$\mu_h(t,x) = h^2(B_1(t,x) + B_2(t,x)) + O\left(h^4\right) \tag{4.15}$$

and

$$\sigma_h^2(t,x) = \frac{1}{nh}\left(s_1^2(t,x) + s_2^2(t,x) + 2s_3^2(t,x)\right) + O(n^{-1}h). \tag{4.16}$$

The expressions $B_1(t,x)$ and $B_2(t,x)$ in (4.15) correspond to the terms $B_1^c(t,x)$ and $B_2^c(t,x)$ in (4.9) – (4.10) after replacing $G(t \mid x)$ with $G_0(t \mid x)$:

$$\begin{aligned} B_1(t,x) =& \frac{1-S(t \mid x)}{2p^2(x)\, m(x)} d_K \left[2\left(1-p(x)\right)' m'(x) + \left(1-p(x)\right)'' m(x)\right] \\ &- \frac{\left(1-p(x)\right)\left(1-S(t \mid x)\right)}{p^2(x)} d_K \int_0^{\tau_0} \frac{G_0'(v^- \mid x)}{1-G_0(v^- \mid x)} \frac{d}{ds}\left(\frac{S'(s \mid x)}{S(s \mid x)}\right)\Bigg|_{s=v^-} dv \end{aligned}$$

and

$$\begin{aligned} B_2(t,x) =& \frac{d_K}{2p(x)\, m(x)}\left(2S'\left(t^- \mid x\right) m'(x) + S''\left(t^- \mid x\right) m(x)\right) \\ &- \frac{S(t \mid x)\, d_K}{p(x)} \int_0^{t} \frac{G_0'(v^- \mid x)}{1-G_0(v^- \mid x)} \frac{d}{ds}\left(\frac{S'(s \mid x)}{S(s \mid x)}\right)\Bigg|_{s=v^-} dv, \end{aligned}$$

where $d_K = \int v^2 K(v)dv$ and $S'(t \mid x), S''(t \mid x)$, $(1-p(x))'$ and $(1-p(x))''$ refer to the derivatives with respect to x.

As the asymptotic bias of the proposed estimator depends on the derivatives of $1-\pi(x)$ and $G_0(t \mid x)$, it is not straightforward to evaluate the exact gain of considering the cure status information in terms of bias.

On the other hand, the functions $s_1(t,x), s_2(t,x)$ and $s_3(t,x)$ are derived from $s_{c,1}(t,x), s_{c,2}(t,x)$ and $s_{c,3}(t,x)$ in (4.11) – (4.13) by replacing $H_{11}(t \mid x)$ with 0. When the cure information is included in the estimation, then $H^{11}(t \mid x) \geq 0$ and therefore $s_{c,i}^2(t,x) \leq s_i^2(t,x)$, for $i = 1,2,3$. So, ignoring the cure status information can increase the variance.

A generalization of the LC estimator in (1.6) in which two different bandwidths are used is

$$\widehat{S}_{0,h_1,h_2}(t \mid x) = \begin{cases} \dfrac{\widehat{S}_{h_2}(t \mid x) - (1-\widehat{p}_{h_1}(x))}{\widehat{p}_{h_1}(x)} & \text{if } 0 \leq t \leq T_{(n)}^1 \text{ and} \\ & \quad \widehat{S}_{h_2}(t \mid x) > 1-\widehat{p}_{h_1}(x) \\ 0 & \text{otherwise.} \end{cases} \tag{4.17}$$

Henceforth, it will be referred to as LC2b. Note that the LC2b estimator has not yet been considered elsewhere.

4.2.3 Bootstrap bandwidth selection

In this section, a bootstrap selection method for the bandwidths h_1 and h_2 of the estimator $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ is proposed. The principle of bootstrap-based bandwidth selection methods is to minimize a bootstrap estimate of the MISE. The bootstrap MISE can be approximated by

$$\mathrm{MISE}^*_x(h_1,h_2) \simeq \frac{1}{B}\sum_{b=1}^{B}\int \left(\widehat{S}^{c,*b}_{0,h_1,h_2}(v \mid x) - \widehat{S}^c_{0,g_{1_x},g_{2_x}}(v \mid x)\right)^2 \omega(v,x)dv, \quad (4.18)$$

where $\widehat{S}^{c,*b}_{0,h_1,h_2}(t \mid x)$ is the estimator computed with the bth bootstrap resample using the bandwidths h_1 and h_2, and $\widehat{S}^c_{0,g_{1_x},g_{2_x}}(t \mid x)$ is the estimator computed with the original sample using the pilot bandwidths g_{1_x} and g_{2_x}. Note that $\omega(v,x)$ is a nonnegative weight function, intended to give lower weight to the right tail of the distribution.

Specifically, the steps to compute the bootstrap bandwidths for a fixed covariate value x are:

Step 1. With the original sample and the pilot bandwidths g_{1_x} and g_{2_x}, compute $\widehat{S}^c_{0,g_{1_x},g_{2_x}}(t \mid x)$ in (4.2).

Step 2. Choose two dense enough grids of bandwidths, $\{h_{11},\ldots,h_{1L}\}$ and $\{h_{21},\ldots,h_{2L}\}$.

Step 3. Generate B bootstrap resamples $\{(X_i^{*(b)},T_i^{*(b)},\delta_i^{*(b)},\xi_i^{*(b)}\nu_i^{*(b)}) : i=1,\ldots,n\}$, for $b=1,\ldots,B$.

Step 4. With the bth bootstrap resample and the bandwidths h_{1j} and h_{2k} compute $\widehat{S}^{c,*b}_{0,h_{1j},h_{2k}}(v \mid x)$, for $j,k=1,\ldots,L$.

Step 5. For h_{1j} and h_{2k}, $j,k=1,\ldots,L$, compute the $\mathrm{MISE}^*_x(h_{1j},h_{2k})$ given by (4.18).

Step 6. The bootstrap bandwidths, h^*_{1x} and h^*_{2x}, are the bandwidths from the grids $\{h_{11},\ldots,h_{1L}\}$ and $\{h_{21},\ldots,h_{2L}\}$ that minimize $\mathrm{MISE}^*_x(h_{1j},h_{2k})$.

The bootstrap resamples in Step 3 are generated following any of the two equivalent resampling algorithms introduced in Section 2.3.3. For computational efficiency (see López-Cheda et al., 2017a,b), we fixed $X^*_i = X_i$ instead of resampling it randomly from $\{X_1,\ldots,X_n\}$. For the pilot bandwidths g_{1x} and g_{2x} the bootstrap bandwidth selectors in Sections 2.3.4 and 3.2.3, respectively, are proposed.

Table 4.1: Characteristics of the simulated scenarios.

Scenario	**Setting**		**% censoring**		**% cured**
	$p(x)$	$S_0(t \mid x)$	$\pi(x) = 0.2$	$\pi(x) = 0.8$	
1	$p_1(x)$	$S_{0,3}(t \mid x)$	52.0	48.2	46.0
2	$p_2(x)$	$S_{0,3}(t \mid x)$	37.5	24.3	20.0
3	$p_3(x)$	$S_{0,3}(t \mid x)$	84.5	81.2	80.0
4	$p_1(x)$	$S_{0,1}(t \mid x)$	51.0	48.0	46.0
5	$p_4(x)$	$S_{0,2}(t)$	45.8	36.5	33.3

A similar bootstrap procedure can be used to select the bandwidth h of the estimator $\widehat{S}^c_{0,h}(t \mid x)$ in (4.3). For choosing the local pilot bandwidths g_{1_x} and g_{2_x}, we propose the same expressions as that in Sections 2.3.4 and 3.2.3.

4.3 Simulation study

A simulation study has been conducted to evaluate the finite sample performance of the estimators $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ in (4.2) and $\widehat{S}^c_{0,h}(t \mid x)$ in (4.3). Particularly, the effect of varying the proportion of individuals identified as cured, $\pi(x)$, the sample size, n, and the form of $p(x)$ and $S_0(t \mid x)$ is studied. The comparisons were made among five estimators: the proposed estimator $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$, its simplified version with a single bandwidth $\widehat{S}^c_{0,h}(t \mid x)$, the LC estimator $\widehat{S}_{0,h}(t \mid x)$, the LC2b estimator $\widehat{S}_{0,h_1,h_2}(t \mid x)$, and the semiparametric latency estimator, $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$, proposed by Bernhardt (2016). The semiparametric estimator includes the information of the identified cured individuals but does not benefit from the flexibility of the nonparametric estimator, its performance strongly depends on the adequacy of the AFT approach to model the latency function.

The proportion of individuals identified as being cured was set to $\pi(x) = 0.2$ and 0.8. The censoring time C was generated, independently of Y given X, so that with probability $\pi(x)$ then $C = \infty$, and with probability $1 - \pi(x)$ the value of C was generated from a Weibull distribution with shape parameter $\alpha = 2$ and scale parameter $\beta = 2$. The covariate X was uniformly distributed on $[-20, 20]$. Four different models for the probability of the event $p(x)$ are assumed: a logistic regression model $p_1(x) = (0.476 + 0.358x) / (1 + \exp(0.476 + 0.358x))$; the second and third models assume the constant functions $p_2(x) = 0.8$ and $p_3(x) = 0.2$; and the last one assumes a quadratic function $p_4(x) = 1 - 0.0025x^2$. Further, three different latency functions were defined as follows. The latency functions $S_{0,1}\left(t \mid x\right)$ in (2.40), and

$$S_{0,2}\left(t\right)=\begin{cases}\dfrac{\exp\left(-t\right)-\exp\left(-4.605\right)}{1-\exp\left(-4.605\right)} & \text{if } 0\leq t\leq 4.605\\ 0 & \text{otherwise.}\end{cases},$$

correspond to the truncated exponential function. The other latency function from an AFT model with lognormal distribution:

$$S_{0,3}\left(t \mid x\right)=\begin{cases}\dfrac{\Psi\left(4.305+0.15x\right)-\Psi\left(\log t-\left(0.3-0.15x\right)\right)}{1-\Psi\left(4.305+0.15x\right)} & \text{if } 0\leq t\leq 4.605\\ 0 & \text{otherwise,}\end{cases}$$

where $\Psi(\cdot)$ is the survival function of the standard normal distribution. This is the parametric model for the latency function that the semiparametric estimator $S_0(t \mid x, \widehat{\boldsymbol{\beta}})$ assumes for the estimation of $S_0(t \mid x)$. Table 4.1 shows the characteristics of the scenarios simulated in our study, which result from combining different models for $p(x)$ and $S_0(t \mid x)$. Scenarios $1-3$ are intended to assess the effect of the form of $p(x)$ on the estimation of the latency function. Specifically, Scenario 1 considers the parametric functions assumed by the semiparametric estimator in Bernhardt (2016), so its performance is expected to be quite good. Note that Scenarios $2-5$ serve to evaluate the performance of the semiparametric estimator when the parametric model assumptions are unsatisfied for either $p(x)$, $S_0(t \mid x)$ or both. Finally, in Scenario 5 the forms of neither the probability of the event $p(x)$ nor the latency $S_0(t \mid x)$ match the models assumed by the semiparametric estimator.

For each scenario, 1000 datasets of sample sizes $n = 50, 100$ and 200 were generated. The search of the optimal bandwidths (h_1, h_2) and h for the nonparametric estimators was performed in grids of 51 values from 5 to 100, equispaced in a logarithmic scale. The Epanechnikov kernel was used. As weight function, $\omega(t, x) = \mathbf{1}(0 \leq t \leq w_x)$ was considered, where w_x is the 90th percentile of $S_0(t \mid x)$.

The integrated squared bias, integrated variance and MISE for each estimator in every scenario were approximated. Figure 4.1 displays the MISE when $n = 100$ and $\pi(x) = 0.8$ for Scenarios $1-5$. As expected, in Scenario 1 the semiparametric estimator behaves well for most values of the covariate X. The semiparametric estimator also gives acceptable results in Scenarios 2 and 3 because of the AFT model in $S_{0,2}(t \mid x)$ and the cure probability function is constant. Nevertheless, in these scenarios, the estimators $\widehat{S}^c_{0,h_1,h_2}\left(t \mid x\right)$ and $\widehat{S}^c_{0,h}\left(t \mid x\right)$ are highly competitive and beat the $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$ for a wide range of values of the covariate. Moreover, the estimators $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ and $\widehat{S}^c_{0,h}(t \mid x)$ outperform $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$ in Scenarios 4–5 because the underlying logistic and/or AFT model assumptions for the cure probability and the latency are violated.

In most scenarios, observe that the $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ and $\widehat{S}^c_{0,h}(t \mid x)$ outperform $\widehat{S}_{0,h_1,h_2}(t \mid x)$ and $\widehat{S}_{0,h}(t \mid x)$. Clearly, the estimator $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ displays a better performance than the simplified version $\widehat{S}^c_{0,h}(t \mid x)$, specially in Scenarios 1, 4 and 5 where the cure probability function is not constant. The conclusion is the same when the cure status information is ignored, the overall better performance of $\widehat{S}_{0,h_1,h_2}(t \mid x)$ over $\widehat{S}_{0,h}(t \mid x)$ implies that the use of two bandwidths is preferable at most times.

Figure 4.2 (top) and Table 4.3 show the effect of the sample size on the behavior of the estimators, comparing the results when $n = 50, 100, 200$ and for $\pi(x) = 0.8$ in Scenario 4. As the sample size increases, the differences between the MISE of all estimators decrease. Table 4.3 lists the effect of the sample size on the MISE and the integrated squared bias and variance. For this scenario, the integrated variances of the estimator $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ are smaller than $\widehat{S}_{0,h_1,h_2}(t \mid x)$ and $\widehat{S}_{0,h}(t \mid x)$ for all sample sizes. These differences decrease as the sample size n increases.

Figure 4.2 (bottom) and Table 4.4 provide some insight about the effect of the cure status $\pi(x)$ on the estimators. The behavior of $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$, $\widehat{S}^c_{0,h}(t \mid x)$ and $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$ generally improves as $\pi(x)$ increases. The estimator $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ substantially outperforms $\widehat{S}_{0,h_1,h_2}(t \mid x)$ and $\widehat{S}_{0,h}(t \mid x)$ in all scenarios.

The performance of the bootstrap bandwidth selector was assessed using a total of 1000 samples, $B = 500$ bootstrap resamples and a grid of 51 bandwidths equispaced in a logarithmic scale, from $h_{i1} = 5$ to $h_{i51} = 100$, for $i = 1, 2$. Figure 4.3 shows the contour plots of the MISE and the density of the selected bootstrap bandwidths $h^*_{1_x}, h^*_{2_x}$ for the covariate values $x = -2, 2, 8$ and 14. For this sample size and this scenario, the bootstrap bandwidth selector seems to perform well.

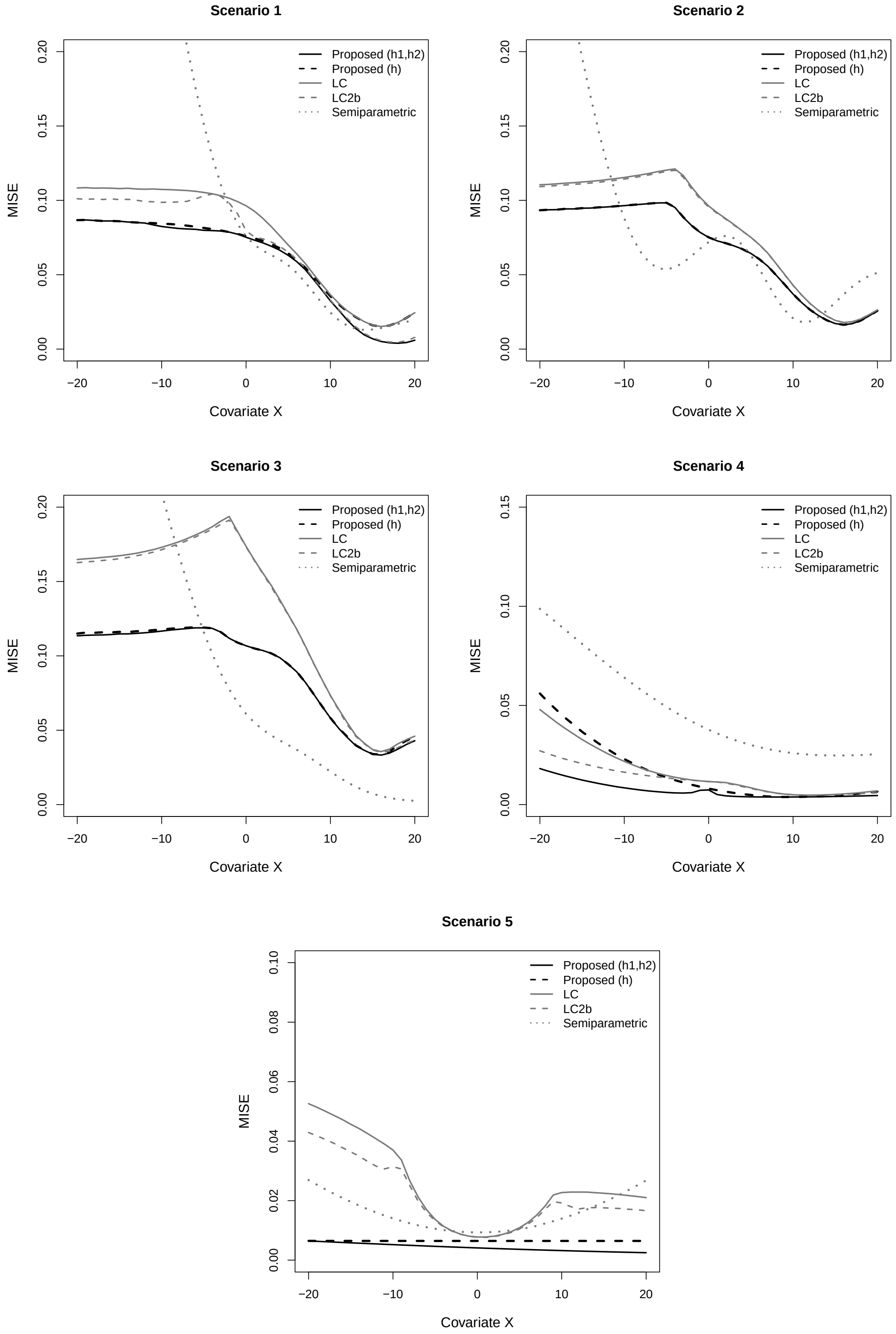

Figure 4.1: MISE of $\widehat{S}^{c}_{0,h_1,h_2}(t \mid x)$, $\widehat{S}^{c}_{0,h}(t \mid x)$, $\widehat{S}_{0,h_1,h_2}(t \mid x)$ $\widehat{S}_{0,h}(t \mid x)$ (all computed with the optimal bandwidth(s)), and $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$ in the simulated scenarios for $\pi(x) = 0.8$ and $n = 100$.

Table 4.2: Integrated squared bias (Ibias^2), integrated variance (Ivar) and MISE of $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$, $\widehat{S}^c_{0,h}(t \mid x)$, $\widehat{S}_{0,h_1,h_2}(t \mid x)$ $\widehat{S}_{0,h}(t \mid x)$ (all computed with the optimal bandwidth(s)), and $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$ in the simulated scenarios for $\pi(x) = 0.8$ and $n = 100$.

		$\widehat{S}^c_{0,h_1,h_2}(t \mid x)$				$\widehat{S}^c_{0,h}(t \mid x)$				$\widehat{S}_{0,h_1,h_2}(t \mid x)$				$\widehat{S}_{0,h}(t \mid x)$				$S_0(t \mid x; \widehat{\boldsymbol{\beta}})$		
Scenarios	x	(h_1, h_2)	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$	h	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$	(h_1, h_2)	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$	h	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$
1	-10	$(14.70, 15.61)$	4.76	3.49	8.24	16.57	5.78	2.66	8.44	$(10.90, 25.21)$	6.26	3.60	9.87	22.36	8.42	2.32	10.74	27.72	2.30	30.02
	0	$(10.26, 8.08)$	5.53	2.00	7.52	8.08	5.72	1.88	7.60	$(13.85, 10.90)$	5.27	2.69	7.96	11.57	7.94	1.70	9.64	6.64	0.94	7.58
	10	$(15.61, 6.75)$	2.47	0.77	3.24	5.99	2.85	0.70	3.55	$(10.90, 25.21)$	2.49	0.80	3.30	5.99	2.94	0.72	3.66	2.18	0.27	2.45
2	-10	$(28.42, 30.17)$	9.05	0.59	9.65	30.17	9.06	0.59	9.65	$(100, 30.71)$	10.88	0.56	11.43	36.11	10.98	0.56	11.54	8.32	0.47	8.79
	0	$(8.07, 8.57)$	6.21	1.30	7.51	8.57	6.21	1.30	7.52	$(100, 8.57)$	8.06	1.49	9.56	8.08	8.16	1.46	9.62	6.82	0.38	7.20
	10	$(6.35.6.75)$	2.93	0.77	3.70	6.75	2.93	0.77	3.71	$(40.71, 7.16)$	3.33	0.96	4.29	5.64	3.24	1.06	4.30	1.49	0.59	2.08
3	-10	$(74.11, 32.03)$	9.29	2.38	11.67	30.17	9.13	2.64	11.77	$(45.89, 13.04)$	14.25	2.89	17.14	100.00	14.27	3.03	17.30	8.32	0.47	8.79
	0	$(13.85, 13.85)$	7.29	3.39	10.68	13.85	7.29	3.39	10.68	$(28.42, 6.75)$	12.84	4.49	17.33	13.04	12.94	4.43	17.37	6.82	0.38	7.20
	10	$(10.26, 10.90)$	3.55	2.27	5.82	10.90	3.55	2.28	5.83	$(45.89, 100)$	4.36	2.96	7.32	10.26	4.41	2.93	7.34	1.49	0.59	2.08
4	-10	$(100, 36.11)$	0.29	0.55	0.85	22.36	1.21	1.09	2.30	$(100, 45.89)$	0.22	1.41	1.64	74.11	0.61	1.57	2.18	3.20	3.21	6.41
	0	$(15.61, 21.06)$	0.15	0.59	0.74	18.68	0.22	0.59	0.81	$(100, 100)$	0.06	1.10	1.16	100	0.06	1.10	1.16	2.76	1.02	3.78
	10	$(58.32, 54.93)$	0.00	0.38	0.39	30.17	0.01	0.38	0.39	$(12.28, 12.28)$	0.03	0.47	0.50	12.28	0.03	0.47	0.50	2.21	0.39	2.60
5	-10	$(83.55, 94.18)$	0.01	0.52	0.52	100	0.00	0.65	0.65	$(10.9, 25.21)$	1.75	1.39	3.14	25.21	2.85	0.86	3.70	0.21	1.20	1.40
	0	$(88.71, 100)$	0.01	0.40	0.41	78.69	0.00	0.65	0.65	$(13.85, 10.90)$	0.14	0.63	0.77	10.26	0.14	0.64	0.78	0.25	0.69	0.93
	10	$(88.71, 100)$	0.01	0.32	0.32	100	0.00	0.65	0.65	$(10.90, 25.21)$	1.12	0.80	1.92	25.21	1.77	0.50	2.27	0.25	1.14	1.40

Table 4.3: Integrated squared bias (Ibias^2), integrated variance (Ivar) and MISE of $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$, $\widehat{S}^c_{0,h}(t \mid x)$, $\widehat{S}_{0,h_1,h_2}(t \mid x)$ $\widehat{S}_{0,h}(t \mid x)$ (all computed with the optimal bandwidth(s)), and $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$ in Scenario 4 for $\pi(x) = 0.8$ and $n = 50, 100, 200$.

		$\widehat{S}^c_{0,h_1,h_2}(t \mid x)$				$\widehat{S}^c_{0,h}(t \mid x)$				$\widehat{S}_{0,h_1,h_2}(t \mid x)$				$\widehat{S}_{0,h}(t \mid x)$				$S_0(t \mid x; \widehat{\boldsymbol{\beta}})$		
n	x	(h_1, h_2)	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$	h	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$	(h_1, h_2)	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$	h	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$
50	-10	$(100, 36.11)$	0.23	1.19	1.42	26.76	1.65	1.56	3.21	$(100, 46.22)$	0.16	2.08	2.24	61.92	0.81	2.15	2.97	3.84	6.42	10.25
	0	$(21.06, 23.74)$	0.31	0.98	1.28	22.36	0.32	0.98	1.30	$(100, 54.93)$	0.04	1.59	1.63	100	0.04	1.59	1.63	3.11	2.24	5.35
	10	$(26.76, 26.76)$	0.01	0.75	0.76	26.76	0.01	0.75	0.76	$(13.04, 13.85)$	0.05	0.86	0.91	13.85	0.03	0.88	0.91	2.20	0.83	3.03
100	-10	$(100, 36.11)$	0.29	0.55	0.85	22.36	1.21	1.09	2.30	$(100, 45.89)$	0.22	1.41	1.64	74.11	0.61	1.57	2.18	3.20	3.21	6.41
	0	$(15.61, 21.06)$	0.15	0.59	0.74	18.68	0.22	0.59	0.81	$(100, 100)$	0.06	1.10	1.16	100	0.06	1.10	1.16	2.76	1.02	3.78
	10	$(58.32, 54.93)$	0.00	0.38	0.39	30.17	0.01	0.38	0.39	$(12.28, 12.28)$	0.03	0.47	0.50	12.28	0.03	0.47	0.50	2.21	0.39	2.60
200	-10	$(38.34, 30.17)$	0.28	0.27	0.55	18.68	0.96	0.77	1.73	$(100, 45.89)$	0.25	0.86	1.11	45.89	0.44	1.12	1.56	3.30	1.66	4.96
	0	$(13.04, 18.68)$	0.08	0.33	0.41	15.61	0.15	0.34	0.49	$(100, 100)$	0.07	0.68	0.74	100	0.07	0.68	0.74	2.80	0.49	3.29
	10	$(23.74, 23.74)$	0.00	0.18	0.18	23.74	0.00	0.18	0.18	$(13.04, 11.57)$	0.01	0.24	0.26	10.9	0.03	0.24	0.26	2.24	0.18	2.42

Table 4.4: Integrated squared bias (Ibias^2), integrated variance (Ivar) and MISE of $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$, $\widehat{S}^c_{0,h}(t \mid x)$, $\widehat{S}_{0,h_1,h_2}(t \mid x)$ $\widehat{S}_{0,h}(t \mid x)$ (all computed with the optimal bandwidth(s)), and $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$ in Scenario 4 for $\pi(x) = 0.2, 0.8$ and $n = 100$.

		$\widehat{S}^c_{0,h_1,h_2}(t \mid x)$				$\widehat{S}^c_{0,h}(t \mid x)$				$\widehat{S}_{0,h_1,h_2}(t \mid x)$				$\widehat{S}_{0,h}(t \mid x)$				$S_0(t \mid x; \widehat{\boldsymbol{\beta}})$		
$\pi(x)$	x	(h_1, h_2)	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$	h	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$	(h_1, h_2)	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$	h	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$	Ibias^2 $\times 10^{-2}$	Ivar $\times 10^{-2}$	MISE $\times 10^{-2}$
0.2	-10	$(100, 36.11)$	0.24	0.83	1.07	23.74	1.39	1.28	2.67	$(100, 38.34)$	0.23	1.13	1.36	28.42	1.09	1.64	2.73	0.56	2.97	3.54
	0	$(17.60, 21.06)$	0.22	0.69	0.91	21.06	0.31	0.63	0.94	$(23.74, 26.76)$	0.15	0.86	1.01	25.21	0.16	0.85	1.01	2.48	0.67	3.15
	10	$(19.84, 19.84)$	0.00	0.43	0.43	19.84	0.00	0.43	0.43	$(13.04, 13.85)$	0.01	0.46	0.47	14.7	0.00	0.47	0.47	4.35	2.04	6.39
0.8	-10	$(100, 36.11)$	0.29	0.55	0.85	22.36	1.21	1.09	2.30	$(100, 45.89)$	0.22	1.41	1.64	74.11	0.61	1.57	2.18	3.20	3.21	6.41
	0	$(15.61, 21.06)$	0.15	0.59	0.74	18.68	0.22	0.59	0.81	$(100, 100)$	0.06	1.10	1.16	100	0.06	1.10	1.16	2.76	1.02	3.78
	10	$(58.32, 54.93)$	0.00	0.38	0.39	30.17	0.01	0.38	0.39	$(12.28, 12.28)$	0.03	0.47	0.50	12.28	0.03	0.47	0.50	2.21	0.39	2.60

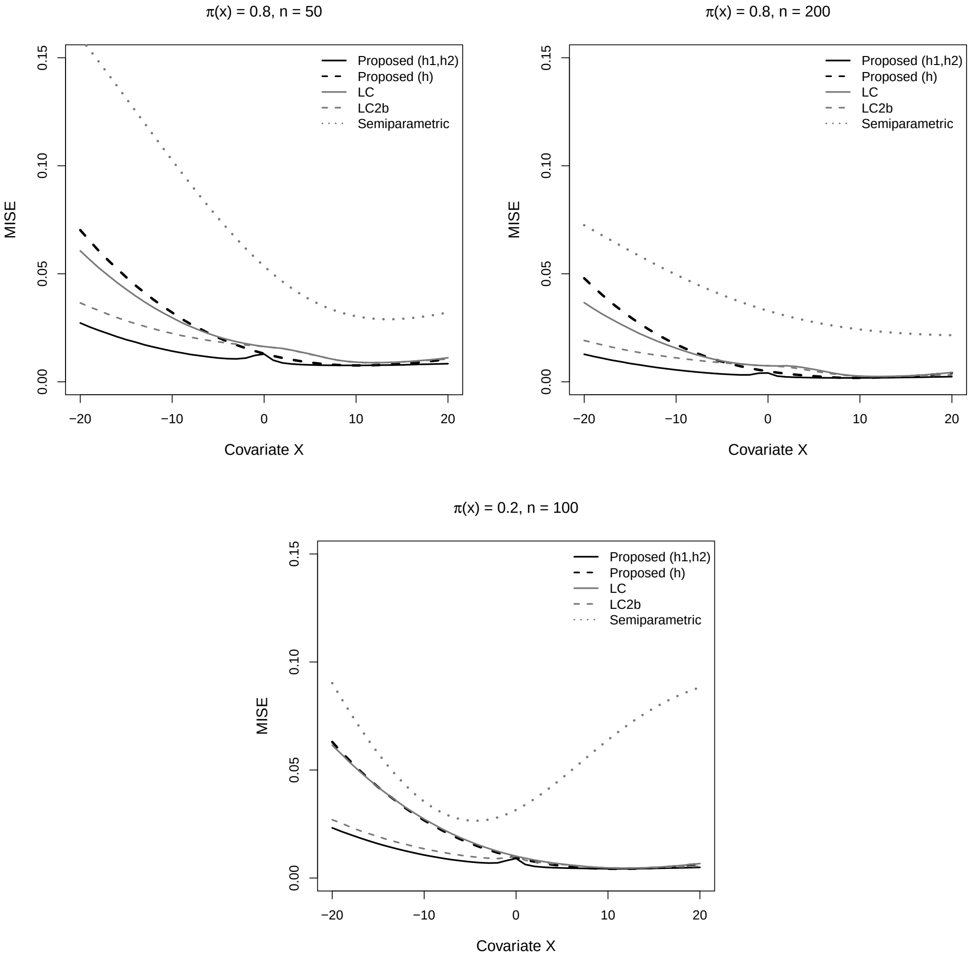

Figure 4.2: MISE of $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$, $\widehat{S}^c_{0,h}(t \mid x)$, $\widehat{S}_{0,h_1,h_2}(t \mid x)$, $\widehat{S}_{0,h}(t \mid x)$ (all computed with the optimal bandwidth(s)), and $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$ in Scenario 4 for $\pi(x) = 0.8, n = 50, 200$ and $\pi(x) = 0.2, n = 100$.

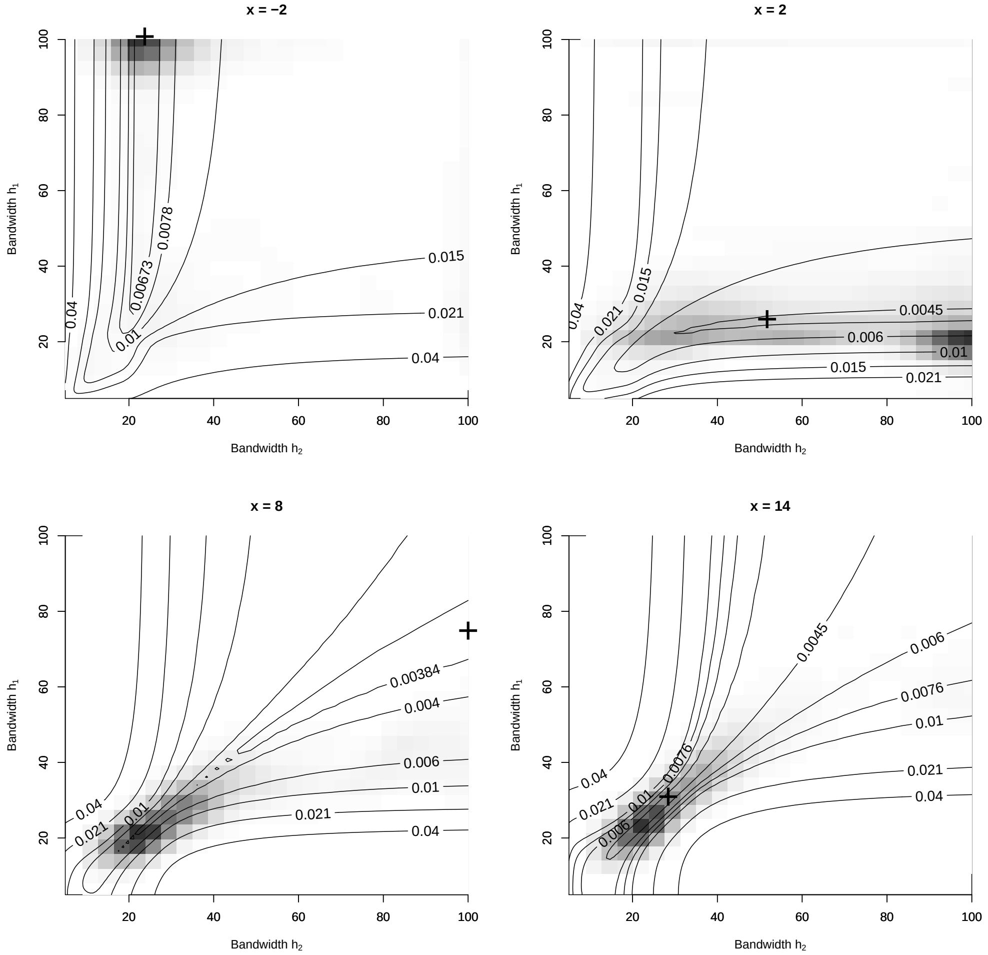

Figure 4.3: Contour plots of the MISE of the proposed estimator in Scenario 4 as a function of the two bandwidths, h_1, h_2. The optimal bandwidth where the minimum MISE is reached is marked with a cross. The joint density of the bootstrap bandwidths $h^*_{1,x}, h^*_{2,x}$ is also shown in shades of gray, where a darker gray represents a greater density.

4.4 Real data analysis

The performance of the estimators $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ in (4.2), $\widehat{S}^c_{0,h}(t \mid x)$ in (4.3) and $\widehat{S}^c_{0,n}(t)$ in (4.4) is illustrated using the COVID-19 data, which was introduced in Sections 1.6 and 3.5.2. Accurate estimates of the trajectory of a patient and the lengths of stay (LoS) from one hospital facility (ward, ICU) to another for proper capacity planning are crucial for healthcare authorities. The total LoS of a patient is divided into several stages according to the patient pathway, and each time can be modeled separately (Currie et al., 2020; García-Vicuña et al., 2021).

For illustration purposes, one specific LoS is considered, the time of a patient in hospital ward until admission to ICU given sex and age as covariates of interest. The latency curves for middle-aged (58 years) and older (74 years) patients were estimated using the estimators $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$, $\widehat{S}^c_{0,h}(t \mid x)$ (both computed with the bootstrap bandwidth selector in Section 4.2.3), $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$, and the estimators $\widehat{S}_{0,h_1,h_2}(t \mid x)$ and $\widehat{S}_{0,h}(t \mid x)$ computed with the `R` package `npcure` (López-Cheda et al., 2021).

The latency function was estimated separately for males and females with $\widehat{S}^c_{0,n}(t)$, the unconditional version of $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$, and the unconditional version of the LC estimator $\widehat{S}_{0,n}(t)$, which ignores a substantial percentage of patients who did not require ICU. Table 3.6 demonstrates the percentage of patients known to be cured from ICU admission is very high, (82.6%). Thus, the non-inclusion of such information in the estimation procedure is expected to give inefficient and possibly biased estimates.

The results are displayed in Figure 4.4. The estimates of the conditional latency functions for the fixed ages show that the survival curves using $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$, $\widehat{S}^c_{0,h}(t \mid x)$ and $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$ are closer to each other, which suggests that the AFT semiparametric model might be appropriate for modeling the LoS until admission to ICU. Meanwhile, the curves estimated using $\widehat{S}_{0,h_1,h_2}(t \mid x)$ and $\widehat{S}_{0,h}(t \mid x)$ appear to be overoptimistic, revealing that ignoring the substantial information of the cured individuals may provide biased estimates. The same conclusion applies to the estimates of the latency function for men and women.

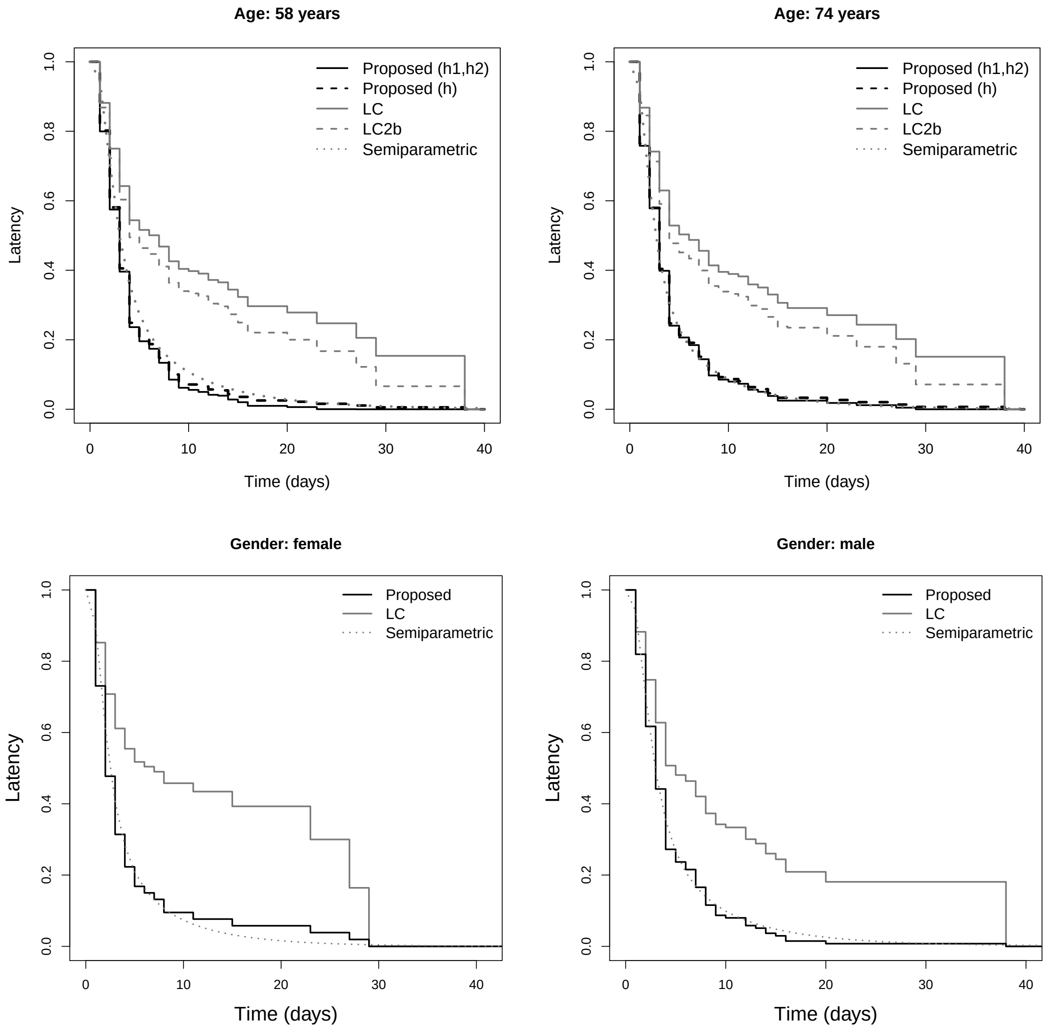

Figure 4.4: Latency estimates of the time in hospital ward until admission to ICU of COVID-19 patients requiring ICU aged 58 and 74 years obtained using $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$, $\widehat{S}^c_{0,h}(t \mid x)$, $\widehat{S}_{0,h}(t \mid x)$, $\widehat{S}_{0,h}(t \mid x)$ (all computed using the bootstrap bandwidths), and $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$. Also shown are the latency estimates by gender, both computed using $\widehat{S}^c_{0,n}(t)$, $\widehat{S}_{0,n}(t)$ and the unconditional version of $S_0(t \mid x; \widehat{\boldsymbol{\beta}})$.

To provide some insight about the source of these differences between the proposed estimator of the latency function that includes the cure status knowledge, and the LC estimator that disregards that available information, Figure 4.5 shows the estimated survival curves for middle-aged (58 years) and older (74 years) patients computed using the estimator $\widehat{S}^c_h(t \mid x)$ and $\widehat{S}_h(t \mid x)$. Also it shows the probabilities of requiring ICU estimated using $1 - \widehat{p}^c_h(x)$ and $1 - \widehat{p}_h(x)$.

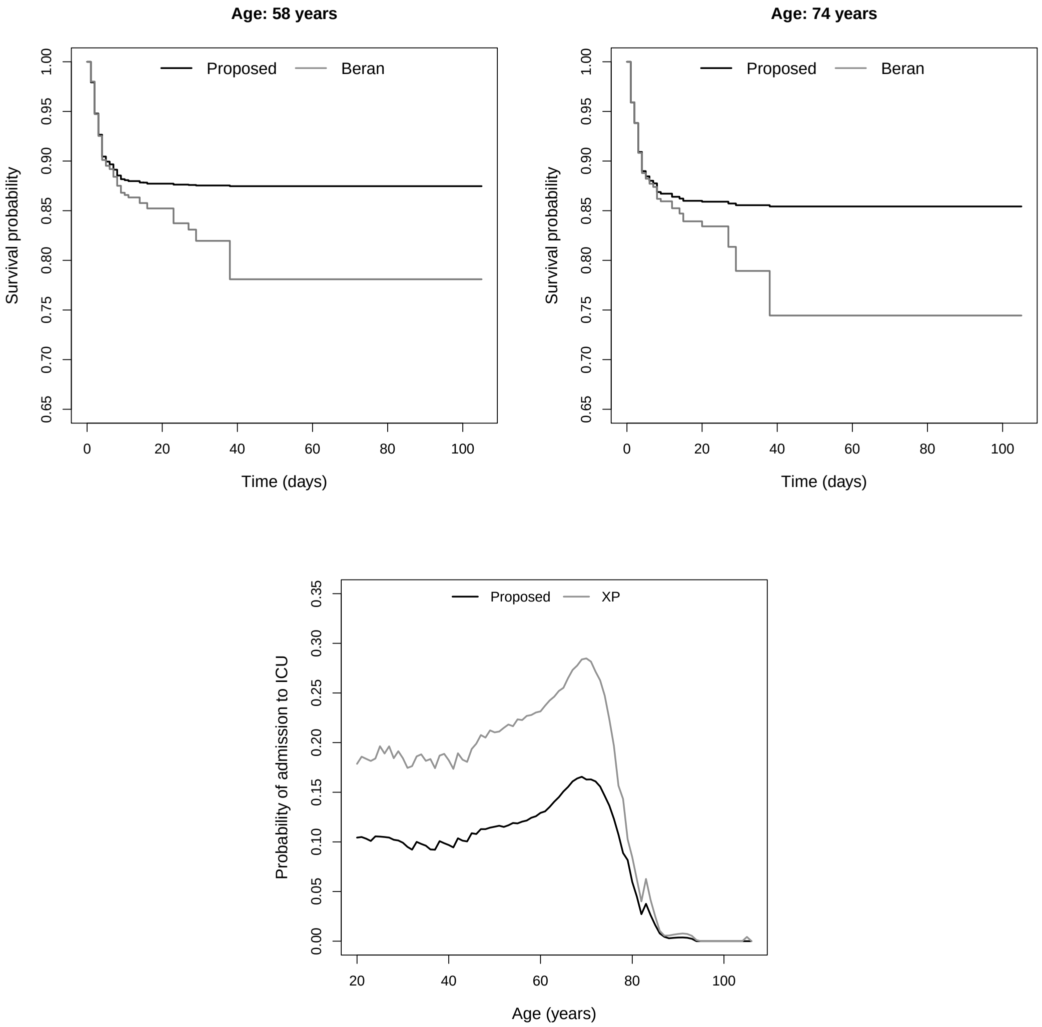

Figure 4.5: (Top panels) Survival estimates of the time in hospital ward until admission to ICU of COVID-19 patients aged 58 and 74 years obtained using $\widehat{S}^c_h(t \mid x)$ and $\widehat{S}_h(t \mid x)$. (Bottom panel) Probability estimates of admission to ICU computed using $1-\widehat{p}^c_h(x)$ and $1-\widehat{p}_h(x)$, all computed with the bootstrap bandwidths.

There are large differences between the estimates of probability of ICU admission obtained with $1-\widehat{p}^c_h(x)$ and $1-\widehat{p}_h(x)$ for patients aged 58 years and only small differences for patients aged 74 years. Meanwhile, the estimators $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ and $\widehat{S}_{0,h}(t \mid x)$ (Figure 4.4) show that the latency estimates are similar when age increases from 58 to 74 years. Notable difference seen in the latency curves obtained with $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ and $\widehat{S}_{0,h}(t \mid x)$ comes from the estimated survival curves in Figure 4.5 (top panels). Survival estimates of the time in hospital ward until admission to ICU obtained with $\widehat{S}^c_h(t \mid x)$ coincide with $\widehat{S}_h(t \mid x)$ at the very beginning. Conversely, as the proportion of patients known to be "cured" is very high, the survival curve estimated with $\widehat{S}_h^c\,(t \mid x)$ tends to level-off quicker than that estimated with $\widehat{S}_h\,(t \mid x)$. This difference is reflected in the latency curves estimated by $\widehat{S}_{0,h_1,h_2}^c\,(t \mid x)$ and $\widehat{S}_{0,h}\,(t \mid x)$, which are simply the estimated survival curves rescaled as proper survival functions.

4.5 Discussion

This chapter proposes a completely nonparametric approach to estimate the latency function in the MCM when the cure status information is available for a subset of the censored observations. The main advantage of the proposed estimator is the flexibility it gives to the procedure since, unlike other alternatives in the literature, it handles any type of covariates without the need of assuming any (semi)parametric model. The asymptotic properties of the estimator were studied, and a procedure for the selection of the bandwidths was provided, showing its good performance in practice.

The effect of including the cure status information in the estimation procedure was demonstrated to be clearly beneficial by comparing the proposed estimator with the nonparametric estimator by López-Cheda et al. (2017b). It has been proved that including the known cures in the estimation reduces the variance asymptotically. The gain in terms of bias is not straightforward from a theoretical point of view, as it depends on the conditional probability of knowing the cure status information and the censoring distribution. Nevertheless, the simulation results show great improvement in terms of bias for all the scenarios.

Chapter 5

A simple nonparametric testing procedure

5.1 Introduction

In general, when analyzing time-to-event data under right censoring, one can never observe the survival time Y and the censoring time C jointly for a given individual. This makes the relation between the survival time and the censoring time unidentifiable. However, a relation between Y and C needs to be assumed to identify the model. Most procedures in survival analysis, including the methods in this thesis, make the key assumption of independent censoring to ensure identifiability. This assumption implies that the mechanism that induces censoring is entirely unrelated to the event of interest. Under independence censoring, these methods used to model survival data by integrating the information coming from both the censored and uncensored observations (see Klein and Moeschberger, 2003; Kalbfleisch and Prentice, 2011, among others).

Independence between Y and C is quite natural and holds very frequently in most contexts. One example is when the censoring mechanism is administrative, given, for instance, by the end of the study, as the loss of information after that date is not related to the survival time. Another example is the loss to follow-up when the reason for the dropout is not related to the survival time (the patient moved to another country, changed contact details, changed the mind and withdraw the study, etc.). Finally, there is also independence when the censoring mechanism is given by a competing risks not related to the event of interest (e.g., the event of interest is death from a given disease and the patient died in a car accident).

Nonetheless, there might be situations where the independence between the survival time and the censoring time is not realistic. In practice, some covariates might be associated to both lifetime and censoring mechanism, inducing dependent censoring. Besides, in medical studies a patient may quit if the health state gets worse because of treatment, so the event (death, relapse) would be close to happen, suggesting a positive correlation between censoring and survival times.

Dependent censoring often hampers analysis in such a way that the estimates obtained using methods that require independence assumption are not valid. Thus, ignoring dependent censoring when it is present typically produces biased estimates. To account for this, Lin et al. (1998), Othus et al. (2009), Ma et al. (2015) and Bernhardt (2016), among others, developed cure models in the presence of dependent censoring. In particular, Bernhardt (2016) proposed a semiparametric mixture cure model that accommodates different censoring distributions for the cured and uncured groups. He also proposed a likelihood ratio test for checking if it is necessary to model different censoring distributions for the cured and uncured individuals. The mentioned tests, however, require a priori assumption on the censoring distribution.

Although independence between survival and censoring times is the most crucial assumption for guaranteeing unbiased inference in survival analysis, it is hardly ever tested. The problem is that it cannot be tested when the data include only the possibly censored observed time and a censoring indicator, as a distribution of Y and C that reconciles the assumption of independence can always be obtained with the observed data (Tsiatis, 1975). So, there is not any formal test statistic to test whether the censoring time is independent of the survival time without assuming further assumptions. The few tests that have been developed rely on extra information that is not typically available to the researcher, or impose equally tenuous auxiliary assumptions. For example, the test of Lee and Wolfe (1998) uses a Cox proportional hazards regression model with a time-dependent covariate and involves further follow-up of a subset of lost-to-follow-up censored subjects. The test of Huang et al. (2004) requires a specific clustered correlation structure among units and imposes independent censoring within each cluster. Frandsen (2019) proposed a test that requires data to include the observed times, censoring times for each observation (censored and uncensored) and covariates, in which, censoring times are assumed to be conditionally independent of the survival times.

In practice, the only way to check the assumption of independence without assuming any further model or extra information is by doing some sensitivity analysis, studying the effect of assuming different degrees of dependent censoring on the parameter estimates (Siannis et al., 2005; Huang and Zhang, 2008; Jackson et al., 2014). In this chapter, a sensitivity analysis of the plausibility of the assumption of independence in the MCM when the cure status is partially known is performed, without requiring any parametric assumption. The procedure relies upon the expected performance of the estimator of the probability of cure proposed in Section 3.2 and the MI-NW estimator described in Section 3.3.2. The motivation comes from the fact that, under the independence assumption, only the estimator proposed in Section 3.2 provides a good approximation of the conditional cure probability. Therefore, large differences between both estimators are expected under independence.

This is an open incomplete study, planned to be finished with a comprehensive simulation study and possibly some theoretical results that might support the proposal.

This chapter is organized as follows. Section 5.2 motivated the simple test proposed to assess the plausibility of the independence assumption. In Section 5.3 a method for implementing the test is introduced. A bootstrap procedure for approximating the null distribution is recommended in Section 5.4.

5.2 Motivation

In the MCM with partially known cures defined in Section 1.5.1, the survival time Y is a random variable with $P\left(Y=\infty \mid X=x\right)=1-p\left(x\right)$ and distribution function for the susceptible individuals $F_0\left(t \mid x\right)=1-S_0(t \mid x)$. Besides, the censoring time C is a random variable with $P\left(C=\infty \mid X=x\right)=\pi\left(x\right)$ and the distribution function for the finite censoring times is $G_0\left(t \mid x\right)$. In this model, the conditional censoring rate is:

$$\begin{aligned}
& P\left(\delta=0 \mid X=x\right) \\
= \; & P\left(C \leq Y \mid X=x\right) \\
= \; & P\left(C \leq Y \mid X=x, Y<\infty\right) P\left(Y<\infty \mid X=x\right) \\
& +P\left(C \leq Y \mid X=x, Y=\infty\right) P\left(Y=\infty \mid X=x\right) \\
= \; & P\left(C \leq Y \mid X=x, Y<\infty, C<\infty\right) P\left(C<\infty \mid X=x, Y<\infty\right)
\end{aligned}$$

$$\begin{aligned}
& \times P(Y<\infty \mid X=x)+P(C \leq Y \mid X=x, Y=\infty) P(Y=\infty \mid X=x) \\
= \; & P(C \leq Y \mid X=x, Y<\infty, C<\infty) P(C<\infty \mid X=x, Y<\infty) p(x) \\
& +(1-p(x)).
\end{aligned}$$

When the independence assumption holds, the conditional censoring rate is

$$\begin{aligned}
P(\delta=0 \mid X=x)=(1-\pi(x)) p(x) \int_0^{\tau_0} S_0(v \mid x) \, d G_0(v \mid x) \\
+(1-p(x)). \qquad (5.1)
\end{aligned}$$

Additionally, the probability of observing the cure status is, for a cured individual

$$P(\xi=1 \mid X=x, Y=\infty)=P(C=\infty \mid X=x, Y=\infty)$$

while for an uncured observation it is

$$\begin{aligned}
& P(\xi=1 \mid X=x, Y<\infty) \\
= & P(Y<C \mid X=x, Y<\infty) \\
= & P(Y<C \mid X=x, Y<\infty, C<\infty) P(C<\infty \mid X=x, Y<\infty) \\
& +P(Y<C \mid X=x, Y<\infty, C=\infty) P(C=\infty \mid X=x, Y<\infty).
\end{aligned}$$

Under conditional independence of the survival time and the censoring time, the probability of observing the cure status is

$$P(\xi=1 \mid X=x, Y=\infty) \;=\; \pi(x)$$

for the cured individuals, and

$$P(\xi=1 \mid X=x, Y<\infty) \;=\; \pi(x)+(1-\pi(x)) \int_0^{\tau_0} S_0(v \mid x) \, d G_0(v \mid x)$$

for the uncured individuals. Therefore, the probability of observing the cure status is different in the cured and uncured subgroups (unless $\int_0^{\tau_0} S_0(v \mid x) \, dG_0(v \mid x) = 0$, which means that all susceptible individuals are censored, so the estimation is not possible). Specifically, the lower $\pi(x)$ and the larger $\int_0^{\tau_0} S_0(v \mid x) \, dG_0(v \mid x)$, the more different the probability of observing the cure status in the cured and uncured individuals. In other words, for a fixed cure rate $1-p(x)$, the higher the conditional censoring rate in (5.1), the more different the conditional probabilities of observing the cure status.

The probability of observing the cure status being equal for the cured and uncured observations is a key assumption for the MI-NW estimator of the cure probability studied in Section 3.3.2:

$$E(\xi \mid X, \nu)=E(\xi \mid X).$$

If that assumption is not met, the MI-NW estimator is clearly biased. Therefore, in the MCM with cures partially known, under conditional independence of Y and C, the MI-NW estimator is expected to perform badly, especially for high levels of censoring, while the proposed estimator $1-\widehat{p}_h^c(x)$ in (3.1) is consistent and asymptotically unbiased. As a consequence, large differences between the proposed cure rate estimator and the MI-NW estimator are expected if the assumption of independence holds, specially when the censoring rate is high.

5.2.1 Sensitivity analysis

In this section, a simulation study is conducted to show the effect of different degrees of dependence in the correlation between Y and C, by comparing the estimator of the cure probability $1-\widehat{p}_h^c(x)$ in (3.1) and the MI-NW estimator $1-\widehat{p}_h^{\text{MI-NW}}(x)$, in (3.19).

The covariate X was generated from a $U[-5,5]$. For generating the survival times Y and the censoring times C under possible dependence, consider a bivariate normal variable

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \quad \text{with } -1 \le \rho \le 1\,.$$

The lifetimes Y were constructed as follows:

$$Y \mid X = \begin{cases} \infty & \text{with probability } 1-p(X) \\ \exp(Z_1) & \text{with probability } p(X) \end{cases}$$

with the cure rate given by

$$1-p(x) = 1 - \frac{\exp(0.476+0.358x)}{1+\exp(0.476+0.358x)}.$$

The censoring times C were simulated to possibly depend on the survival time Y as follows.

$$\text{If } \; Y<\infty \; \text{ then } \begin{cases} C=\infty & \text{with probability } \pi(X)(1-\rho)\mathbf{1}(\rho\ge 0) \\ & \qquad\qquad +\left(\pi(X)-(1-\pi(X))\rho\right) \\ & \qquad\qquad \times\mathbf{1}(\rho<0)\,, \\ C=\exp(\mu_C+\sigma_C Z_2) & \text{otherwise.} \end{cases}$$

$$\text{If } \; Y=\infty \; \text{ then } \begin{cases} C=\infty & \text{with probability } \; (\pi\,(X)+\rho\,(1-\pi\,(X))) \\ & \qquad\qquad\qquad \times\mathbf{1}\,(\rho\geq 0)+\pi\,(X) \\ & \qquad\qquad\qquad \times\,(\rho<0)\,\mathbf{1}\,(1+\rho)\,, \\ C=\exp\left(\mu_C+\sigma_C Z_2\right) & \text{otherwise.} \end{cases}$$

After simulating each pair of times (Y_i, C_i) for $i = 1, \ldots, n$ the observation $(T_i, \delta_i, \xi_i\nu_i)$ is constructed as

$$\begin{aligned} T_i &= \min\left(Y_i, C_i\right) \mathbf{1}(\xi_i\nu_i = 0) + \exp(\mu_C + \sigma_C Z_{2i})\mathbf{1}(\xi_i\nu_i = 1), \\ \delta_i &= \mathbf{1}\left(Y_i < C_i\right), \\ \xi_i\nu_i &= \mathbf{1}(Y_i = \infty, C_i = \infty). \end{aligned}$$

The parameter ρ controls the degree of correlation between the survival times Y and the censoring times C. The value $\rho = 0$ corresponds to the independence assumption, with $P(C = \infty \mid X) = \pi\,(X)$. Positive values of ρ represent positive correlation; when $Y = \infty$ then $P(C = \infty \mid X)$ increases to $\pi\,(X) + \rho\,(1 - \pi\,(X))$, while when $Y < \infty$ then $P(C = \infty \mid X)$ decreases to $\pi\,(X)\,(1 - \rho)$. Likewise, negative values of ρ represent negative correlation.

Under independent censoring, the censoring rate in (5.1) in this simulated scenario is

$$P\,(\delta = 0 \mid X = x) = (1 - \pi\,(x))\,p\,(x)\,(1 - E\,(\Phi\,(\mu_C + \sigma_C Z_2))) + (1 - p\,(x))\,.$$

For fixed x, the level of censoring increases as $\pi(x)$ and μ_C decrease. Finally, σ_C controls the dispersion in the censoring times.

The estimator $1 - \widehat{p}_h^{\mathrm{c}}\,(x)$ is compared with $1 - \widehat{p}_h^{\text{MI-NW}}(x)$ when estimating the cure probability for $x = -2$, whose value is $1 - p\,(-2) = 0.56$. Different settings were considered by varying the values of $\pi\,(x)$, μ_C and ρ. For each simulation setting, 100 datasets are generated for the sample size $n = 100$. The same bandwidth, $h = 3.981$ was used for both estimators. For the pilot bandwidths required for the MI-NW estimator, the same bandwidths $g_1 = g_2 = 3.981$ were considered.

The censoring rate, controlled by $\pi(x)$ and μ_C, plays a key role in the performance of both estimators. The effect of $\pi\,(x)$ on the behavior of $1 - \widehat{p}_h^{\mathrm{c}}(x)$ and $1 - \widehat{p}_h^{\text{MI-NW}}(x)$ is shown in Figure 5.1 under independence ($\rho = 0$), and in Figure 5.2 under strong positive dependence ($\rho = 0.9$). Under independence censoring, $1 - \widehat{p}_h^{\mathrm{c}}\,(x)$ is unbiased giving a good estimates even for high censoring rates (small values of $\pi\,(x)$). On the other hand, $1 - \widehat{p}_h^{\text{MI-NW}}(x)$ is clearly biased when the censoring rate is large (small to moderate values of $\pi(x)$), although the estimates improve as $\pi(x)$ approaches 0.9. Under strong positive dependence, both estimators perform similarly, producing biased estimates regardless of the value of $\pi(x)$. The effect of the censoring rate is given by μ_C, for $\pi(x) = 0.5$, is shown in Figure 5.3 under independence ($\rho = 0$) and in Figure 5.4 under strong positive dependence ($\rho = 0.9$). The estimates are different under independence and quite similar under positive dependence. Finally, the effect of the dependence between Y and C, is given by the value of ρ. As shown in Figure 5.5 when the censoring rate is high ($\mu_C = 0$, $\pi(x) = 0.2$), and in Figure 5.6 when the censoring rate is low ($\mu_C = 0$, $\pi(x) = 0.5$).

Under independence censoring, $1 - \widehat{p}_h^{\mathrm{c}}(x)$ is asymptotically unbiased. Note that when the censoring rate is low, the MI-NW estimator can perform either similarly (when $\pi(x)$ is very high, so the siMAR assumption is close to be acceptable and the MI-NW estimator is also asymptotically unbiased), or quite differently (when μ_C is large, $\int_0^{\tau_0} S_0(v \mid x)\, dG_0(v \mid x)$ is high and the siMAR assumption is very unlikely, so the MI-NW estimator is asymptotically biased). Consequently, testing the plausibility of the independence assumption by comparing both estimators is not straightforward when the censoring rate is low.

However, when the censoring rate is high (low values of $\pi(x)$ or μ_C) the siMAR assumption is far of being fulfilled, and $1 - \widehat{p}_h^{\text{MI-NW}}(x)$ is expected to be clearly biased. In this case, the conclusions are a little bit more direct, under independence both estimators are quite different (see results for $\pi(x) = 0.1$ to 0.4 in Figure 5.1) while they are quite similar under dependence (see Figure 5.2). This fact can be used to introduce a test to be applied for high levels of censoring, by comparing the estimates $1 - \widehat{p}_h^{\mathrm{c}}(x)$ and $1 - \widehat{p}_h^{\text{MI-NW}}(x)$.

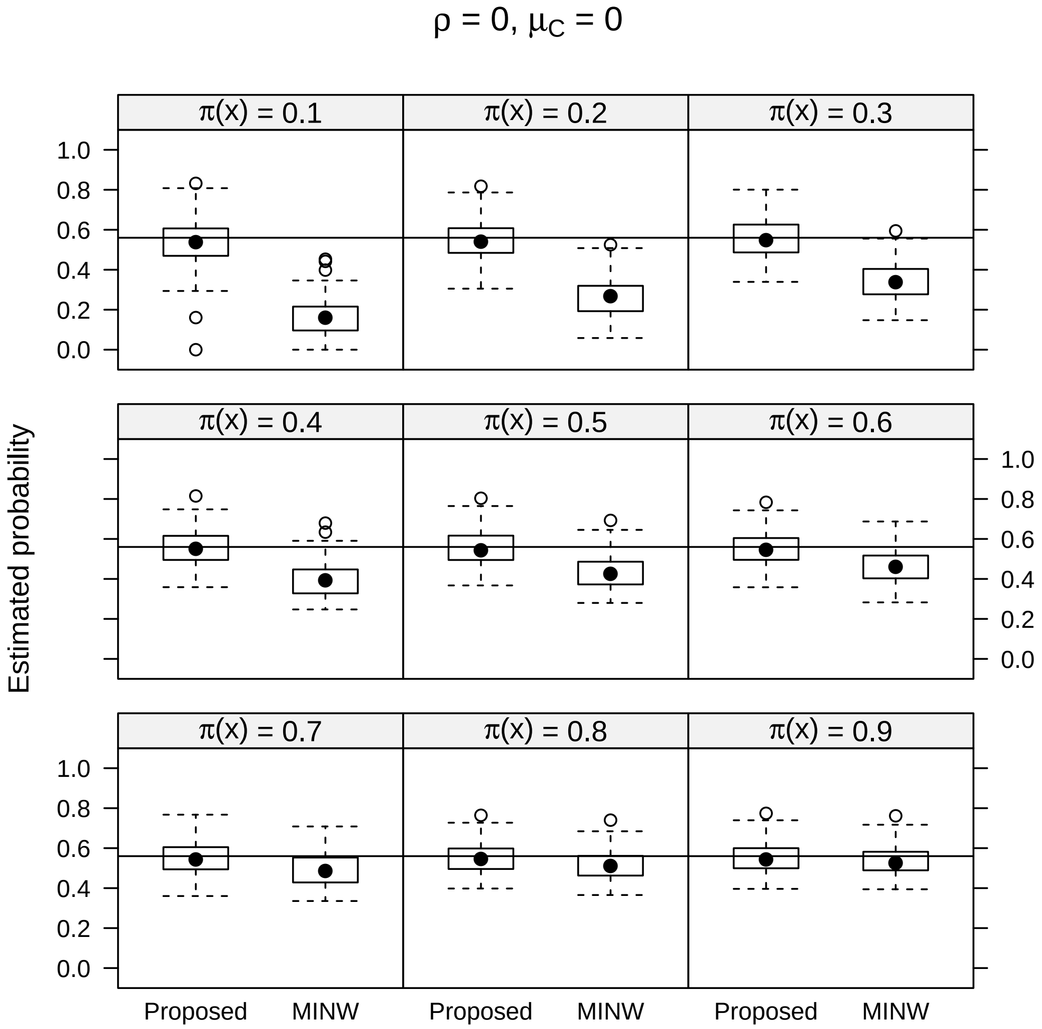

Figure 5.1: Estimation of $1-p\,(-2) = 0.56$ (black horizontal line) using $1-\widehat{p}_h^c(x)$ (Proposed) and $1-\widehat{p}_h^{\text{MI-NW}}(x)$ (MINW), computed with $m = 100$ samples of size $n = 100$ and the same bandwidth $h = 3.981$, for different values of $\pi\,(x)$, when $\mu_C = 0$ and under independence ($\rho = 0$).

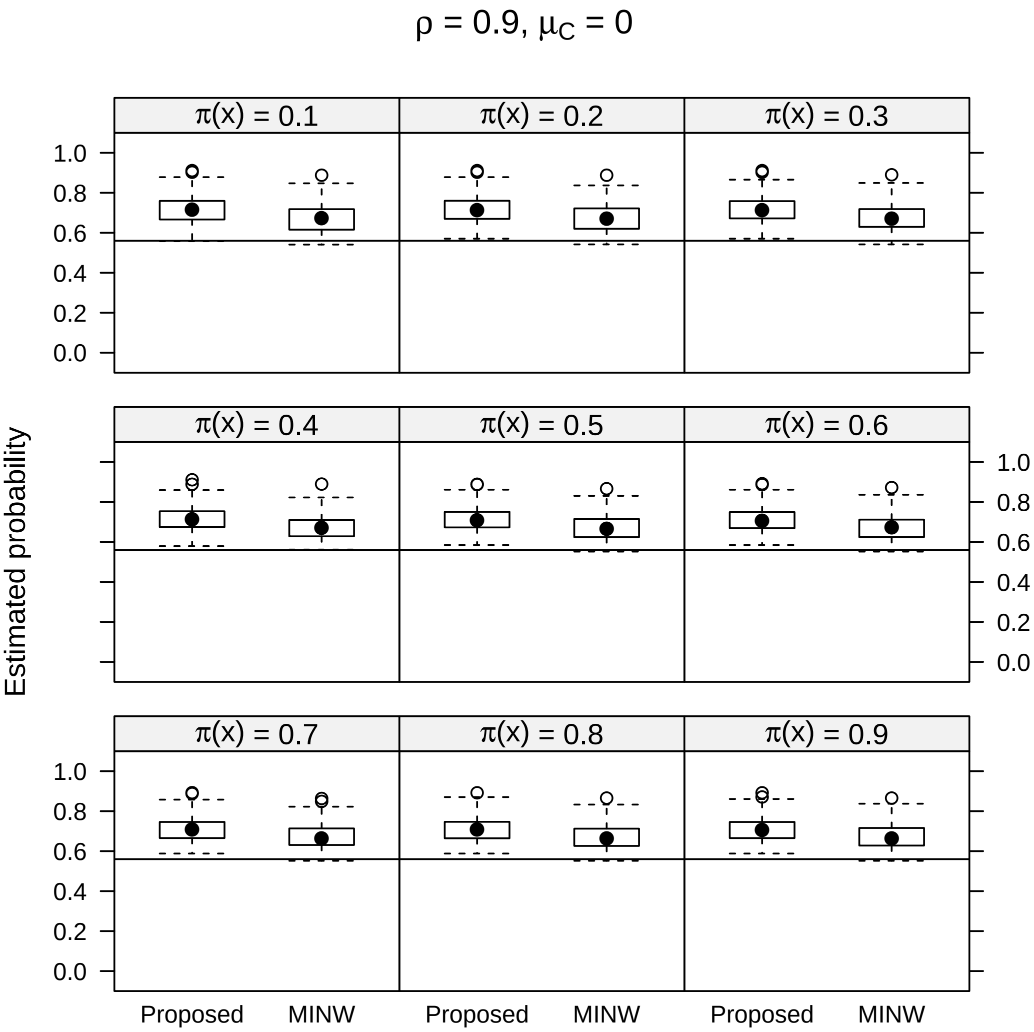

Figure 5.2: Estimation of $1-p\,(-2)=0.56$ (black horizontal line) using $1-\widehat{p}_h^c(x)$ (Proposed) and $1-\widehat{p}_h^{\text{MI-NW}}(x)$ (MINW), computed with $m=100$ samples of size $n=100$ and the same bandwidth $h=3.981$, for different values of $\pi\,(x)$, when $\mu_C=0$ and under strong positive dependence ($\rho=0.9$).

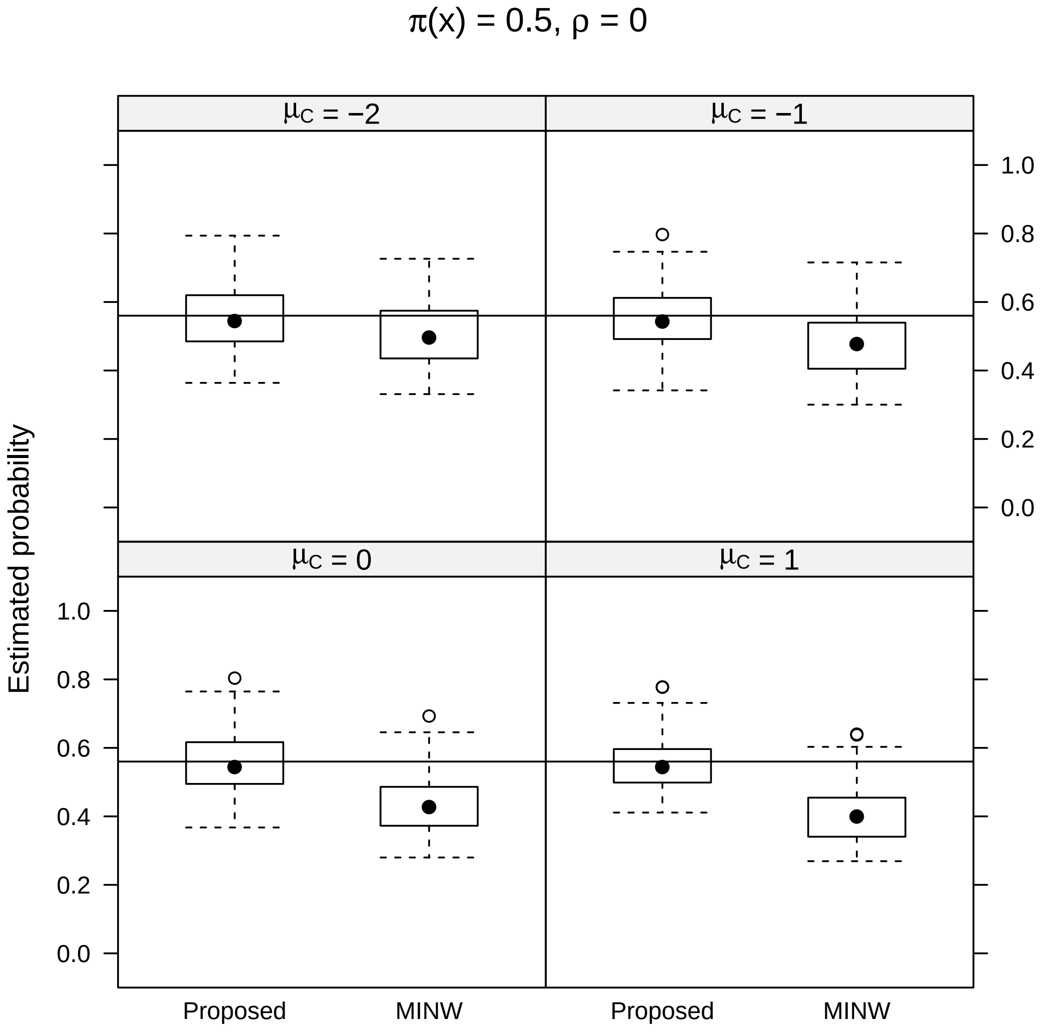

Figure 5.3: Effect of the rate of censoring, given by different values of μ_C, in the estimation of $1-p\,(-2)=0.56$ (black horizontal line) using $1-\widehat{p}_h^c(x)$ (Proposed) and $1-\widehat{p}_h^{\text{MI-NW}}(x)$ (MINW), computed with $m=100$ samples of size $n=100$ and the same bandwidth $h=3.981$, when $\pi\,(x)=0.5$ and under independence $(\rho=0)$.

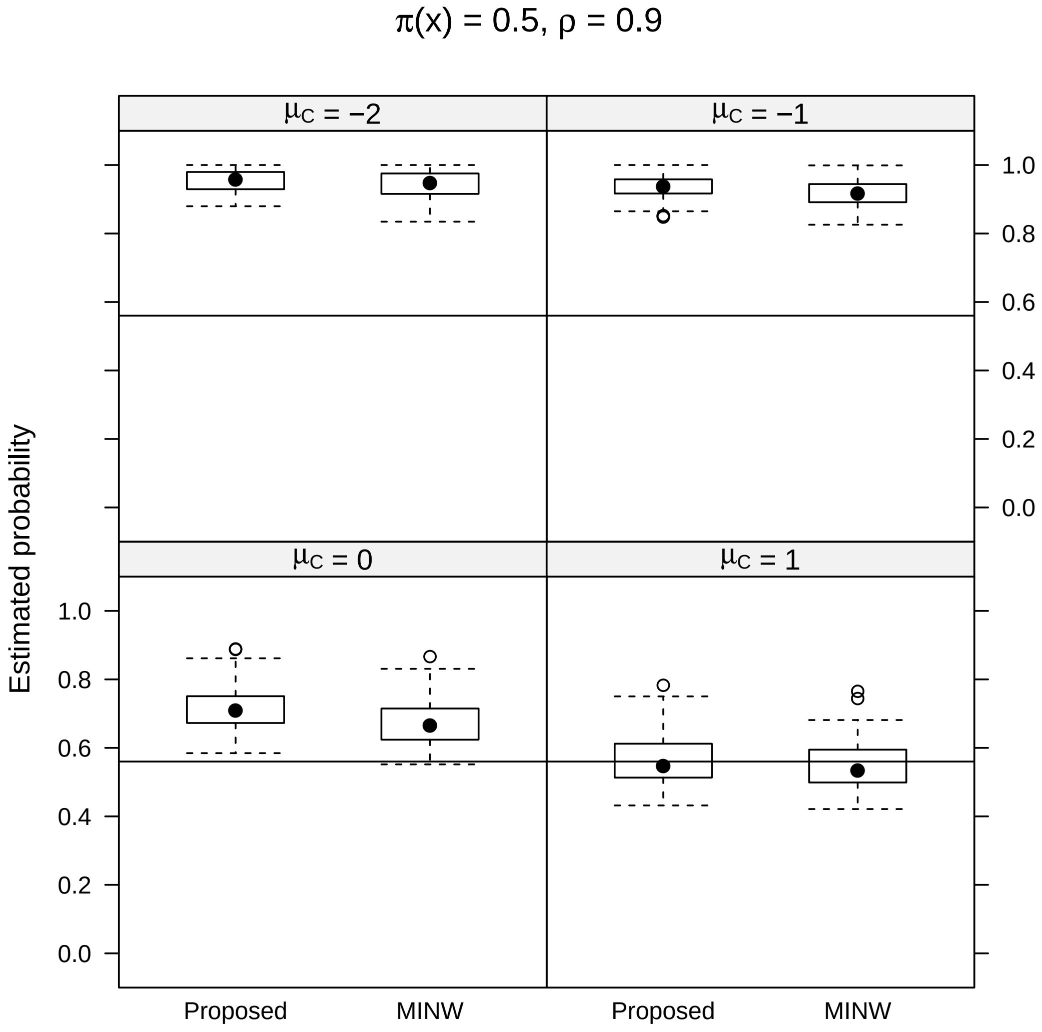

Figure 5.4: Effect of the rate of censoring, given by different values of μ_C, in the estimation of $1-p\,(-2)=0.56$ (black horizontal line) using $1-\widehat{p}_h^c(x)$ (Proposed) and $1-\widehat{p}_h^{\text{MI-NW}}(x)$ (MINW), computed with $m=100$ samples of size $n=100$ and the same bandwidth $h=3.981$, under strong dependence ($\rho=0.9$) when $\pi\,(x)=0.5$.

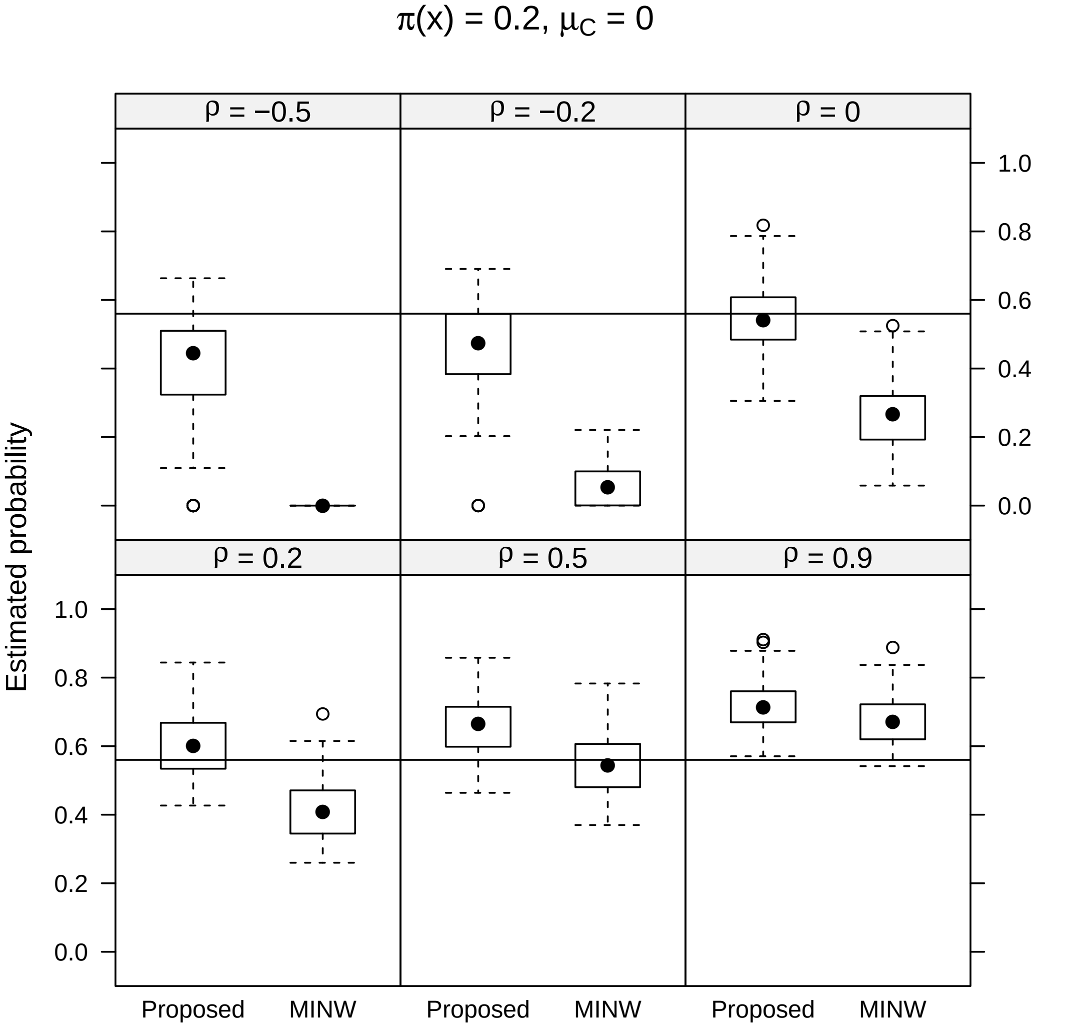

Figure 5.5: Effect of the degree of dependence given by different values of ρ, in the estimation of $1-p\,(-2)=0.56$ (black horizontal line) using $1-\widehat{p}_h^c(x)$ (Proposed) and $1-\widehat{p}_h^{\text{MI-NW}}(x)$ (MINW), computed with $m=100$ samples of size $n=100$ and the same bandwidth $h=3.981$, when $\pi\,(x)=0.2$ and $\mu_C=0$.

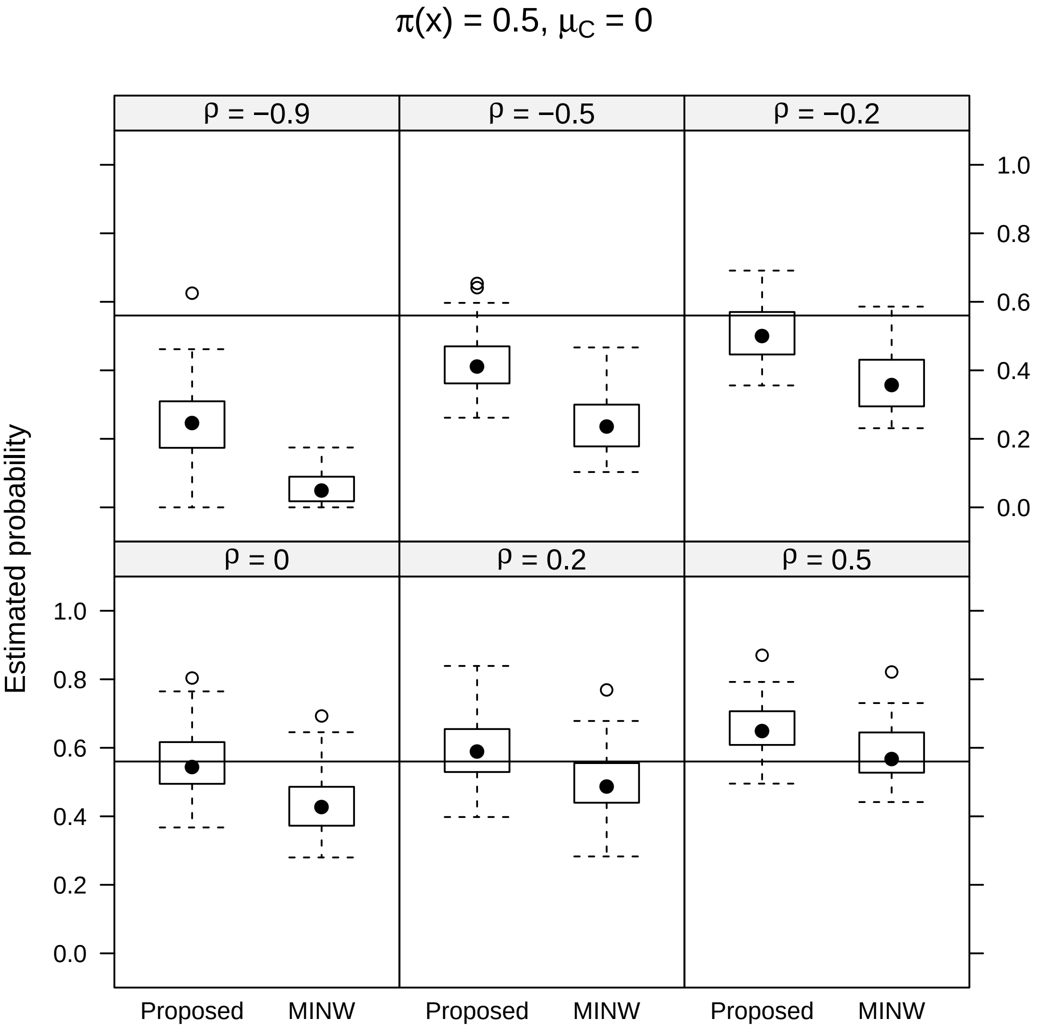

Figure 5.6: Effect of the degree of dependence given by different values of ρ, in the estimation of $1-p\,(-2)=0.56$ (black horizontal line) using $1-\widehat{p}_h^{\mathrm{c}}(x)$ (Proposed) and $1-\widehat{p}_h^{\text{MI-NW}}(x)$ (MINW), computed with $m=100$ samples of size $n=100$ and the same bandwidth $h=3.981$, when $\pi\,(x)=0.5$ and $\mu_C=0$.

5.3 Proposed test statistic

The performance of the MI-NW estimator could be considered as a measure of how plausible the siMAR assumption is, and consequently, of how unlikely the assumption of independence between Y and C is, by computing the distance between $1-\widehat{p}_h^{\mathrm{c}}(x)$ in (3.1) with bandwidth h, and $1-\widehat{p}_{h_3}^{\text{MI-NW}}(x)$ in (3.19) using a bandwidth h_3, possibly different to the bandwidth for $1-\widehat{p}_h^{\mathrm{c}}(x)$, and the pilot bandwidths (g_1,g_2). When the censoring rate is high, if the difference between $1-\widehat{p}_h^{\mathrm{c}}\,(x)$ and $1-\widehat{p}_{h_3}^{\text{MI-NW}}(x)$ is

large there is not evidence against the independence between Y and C, while a very small difference between $1-\widehat{p}_h^c(x)$ and $1-\widehat{p}_{h_3}^{\text{MI-NW}}(x)$ would suggest that the siMAR assumption is plausible, and that there is not independence between Y and C.
The test for checking H_0, that is Y and C are conditionally independent, is based on the difference between the proposed estimator for the cure rate and the MI-NW estimator:

$$T_{h,h_3}(x) = nh^{1/2}\left(\widehat{p}_h^c(x) - \widehat{p}_{h_3}^{\text{MI-NW}}(x)\right)^2. \tag{5.2}$$

The asymptotic distribution of the test statistic in (5.2) will be derived (future research) under the null hypothesis. But it is expected that the test based on the asymptotic distribution does not work well in practice, due to the slow convergence rate. On the contrary, the bootstrap method has been shown in the literature to give accurate estimations of the level in hypothesis tests. In the next section, a bootstrap procedure is proposed to approximate the critical values of the test.

5.4 Bootstrap approximation of the null distribution

The following bootstrap procedure is proposed in order to approximate the critical values of the test in (5.2). Let B be the number of bootstrap samples, the proposed bootstrap procedure consists of the following steps.

Step 1. Simulate the bootstrap sample $\left\{\left(X_i^{*,b}, T_i^{*,b}, \delta_i^{*,b}, \xi_i^{*,b}\nu_i^{*,b}\right), i=1,\dots,n\right\}$, for $b=1,\dots,B$, using the resampling methods of Section 2.3.3.

Step 2. For $b=1,\dots,B$, use the bootstrap sample $\left\{\left(X_i^{*,b}, T_i^{*,b}, \delta_i^{*,b}, \xi_i^{*,b}\nu_i^{*,b}\right), i=1,\dots,n\right\}$, and the bandwidths h for $1-\widehat{p}_h^c(x)$ and (g_1, g_2, h_3) for $1-\widehat{p}_{h_3}^{\text{MI-NW}}(x)$, to compute the test statistic $T_{h,h_3}^{*b}(x)$ in (5.2).

Step 3. Order $\{T_{h,h_3}^{*1}(x), \dots, T_{h,h_3}^{*B}(x)\}$ and select the $[\alpha B]$th order statistic as the critical value, where α is the level of the test. Reject H_0 with level α if $T_{h,h_3}(x)$ is smaller than the critical value.

In a scenario without a cured fraction, Li and Datta (2001) considered the analogous obvious bootstrap algorithm with a single pilot bandwidth $h_0 = h_{0F} = h_{0G}$ for bootstrapping the Beran estimate of the conditional survival function $S(t \mid x)$ and the censoring distribution $G(t \mid x)$. They proved that the optimal order for the pilot h_0

is $n^{-\beta}$ with $\frac{1}{10} < \beta < \frac{1}{5}$, so h_0 goes to zero at a slower rate than the optimal bandwidth for Beran's estimator, whose optimal order is $n^{-1/5}$. Specifically, according to Li and Datta (2001), the order for the pilot bandwidth h_0 that showed satisfactory results is $h_0 \sim n^{-0.11}$. The choice of the pilot bandwidths deserves a thorough analysis, but in practice, in accordance with Li and Datta (2001), one single pilot bandwidth $h_0 = c_h n^{-0.11} = h \times n^{0.09}$ is proposed where $h = c_h n^{-1/5}$ is the bandwidth used in the estimation of the proposed estimator $1 - \hat{p}_h^c(x)$ in the computation of the test statistic.

5.5 Bandwidth selection

Anderson et al. (1994) acknowledged that the performance of kernel-based tests is affected by the selection of the bandwidth, and noticed that the bandwidth must be constant for the test to perform well in terms of power and significance level. Still, methods for bandwidth choice in testing problems have received relatively little attention in literature.

Smoothing-based statistics are a very natural way of testing the goodness-of-fit of the data to a given model specification. However, it is recognized that the choice of the bandwidth may greatly influence the final shape of a smooth estimator, while having a big impact in testing for significance too. The goal in testing problems is different in nature to approximating the underlying population curve, since one will be interested in the construction of a powerful test statistic rather than a good estimator. Optimal data-driven smoothing selectors in the sense of integrated deviations from the target function may not be appropriated in testing problems.

Cao and Van Keilegom (2006) proposed to select the bandwidth that maximizes the power by means of a double-bootstrap procedure; Lindsay et al. (2014) offer a method for selecting the bandwidth that maximizes the power of the test, (see also Gao and Gijbels, 2008; Martínez-Camblor and de Uña-Álvarez, 2013, among others)

Here the suggestion is that the bandwidth h for computing $1 - \widehat{p}_h^c(x)$ and the bandwidth h_3 and (g_1, g_2), for computing $1 - \widehat{p}_{h_3}^{MI-NW}(x)$, are proposed to coincide. The recommendation is to consider a set of suitable values for the bandwidth $h = g_1 = g_2 = h_3 = c_h n^{-1/5}$ and proceed from there.

Chapter 6

Conclusions and future work

This thesis encompasses the attention-raising cure models in survival analysis, when the cure status, usually unknown in standard MCM due to censoring, is partially known for some censored observations. With covariates, this situation has been addressed in the literature only from a parametric or semiparametric point of view.

Theoretical results provided contribute to the field of MCM from a nonparametric approach, illustrate the potential of the proposed methodology, and provide encouraging perspectives for enhancement in MCM when some censored individuals can be classified as cured from the event. The simulations confirm these results under different settings. In particular, when the proportion of knowing cured observations is high the proposed estimators show better results. Applications to three medical data complete the thesis.

As most methods developed in survival analysis, the proposed estimators in Chapters 2 – 4 require the conditional independence between the survival and censoring times given the covariates. This assumption is quite natural and holds in most practical studies. However, there could be situations where lifetimes might be correlated to the censoring distribution. The lack of a formal test for independence without requiring further conditions makes that this assumption is hardly ever tested in practice. In Chapter 5, a simple idea is proposed to assess how plausible the independence assumption is when some individuals are identified to be cured, and the censoring rate is high. Future work related to this point encompasses an extensive simulation study to check the performance of the proposed testing procedure. In addition, the bootstrap bandwidth selection method will be proposed and studied through a simulation study.

The proposed methods are tailored for ordinary right censored time-to-event data. Nonetheless, observations may suffer from other types of censoring, or even from truncation. Approaches that handle these complexities merit further investigation.

In many cases researchers have sought to go beyond a single continuous covariate to multiple covariates. Future exploration is of interest in case of a MCM with cured individuals randomly observed, in presence of multiple covariates. Let $\boldsymbol{X}$ be a d-dimensional vector $(X_1, \ldots, X_d)$ of mixed discrete, categorical and/or continuous variables. Note that when there are many covariates, the sparseness of data gives rise to the well known "curse of dimensionality", which implies that massive amounts of data will be required for accurate estimate as the number of covariates d increases. Different approaches are available in the literature, that enable handling multiple covariates. We briefly discuss both approaches in turn.

In the first approach, the vector of covariates $\boldsymbol{X}$ can be handled using a multivariate kernel function $\boldsymbol{K}$ defined on $\mathbb{R}^d$ depending on a bandwidth parameter $h_n \to 0$ (Liang et al., 2012). Li and Racine (2008) introduced a nonparametric kernel estimator to estimate the joint multivariate cumulative distribution function of mixed discrete and continuous variables, using a generalized multivariate kernel defined as the product of univariate kernels. The multivariate kernel to be used in the estimation is defined as $\boldsymbol{K_h}(\boldsymbol{x}) = \prod_{k=1}^{d} K_{h_k}(x_k - X_k)$, where $K_{h_k}(.)$ is a univariate kernel computed with the univariate covariate X_k and the corresponding optimal bandwidth h_k. If one choose to use this type of multivariate kernel, then an optimal bandwidth must be chosen for each univariate component of the covariate $\boldsymbol{X}$.

Alternative approach is to restrict the form of the effect of the covariates to a *single-index*. This approach is particularly attractive since the original multidimensional covariate vector $\boldsymbol{X}$ is replaced by a 1-dimensional single-index (the linear combination $\boldsymbol{\beta^t X}$). The coefficients $\boldsymbol{\beta}$ characterize the relative importance of $\boldsymbol{X}$. Contrarily to the product kernel, this procedure does not suffer from the curse of dimensionality problems as it summarizes the covariate vector $\boldsymbol{X}$ into a single index. Single-index models are well-studied in the survival analysis literature (Wang et al., 2007; López, 2009; López et al., 2013). The model expresses the covariate vector $\boldsymbol{X}$ as a function of a linear combination of univariate covariates $g(\boldsymbol{\beta}^t \boldsymbol{X})$. Amico et al. (2019) proposed a kernel estimator, based on the NW weights, for estimating $g(.)$ in the context of MCM. The ideas of considering multiple covariates using the single-index (Amico et al., 2019) can be extended to our methodology.

The problem of goodness-of-fit test to assess the aptness of the assumed parametric form was recently studied in the standard MCM. Müller and Van Keilegom (2019) proposed a test to study whether the cure rate, as a function of the covariates, satisfies a certain parametric model. Meanwhile, Geerdens et al. (2020) proposed a test procedure to assess the parametric form imposed on the latency function. Both proposals ignore the cure status information. Therefore, a possible future work would consist in extending these proposals to include the cure status information that these methods ignore.

The development of a variable-selection procedure in the MCM when the cure status is partially known is also of interest. Covariate significance tests for the survival function, cure probability and latency function will be proposed, based on the same ideas in which the covariate significance tests for the cure probability proposed in López-Cheda et al. (2020).

On some occasions, the chance of observing the primary event of interest can be altered or precluded by the occurrence of other events. Caution is needed in estimating the probability of the event of interest (or cure) in the presence of these so-called competing risks. Extending the proposed methodology to accommodate competing events is important because treating them as censored observations could bias the final estimates. Considering the breast cancer example, deaths unrelated to cancer become competing risks events and worth modeling. Multi-state models generalize competing risks model by also describing transitions to intermediate events. Only few authors (Wang, 2003; Conlon et al., 2014; Beesley and Taylor, 2019; Nicolaie et al., 2019) have considered analysis of competing risks or multi-state events with a cure fraction.

The `R` package `npcure` by López-Cheda et al. (2021) provides the nonparametric estimation and testing procedures in MCM proposed by López-Cheda et al. (2017a,b, 2020), including the Beran estimator. The situation when the cure status is partially known is not currently supported by the package but will be considered in future versions. Further, the estimator of the conditional survival function introduced in this paper and subsequent estimators of the cure rate and latency functions will be incorporated in the upgraded package.

Appendices

Appendix A

Proofs of the results in Chapter 2

Motivation of the proposed estimator $\widehat{S}^c_n(t)$.

In an unconditional setting, $\widehat{S}^c_n(t)$ is

$$\widehat{S}^c_n(t) = \prod_{i=1}^n \left(1 - \frac{\delta_{[i]}\mathbf{1}\left(T_{(i)} \leq t\right)}{n - i + 1 + \sum_{j=1}^{i-1}\mathbf{1}\left(\xi_{[j]}\nu_{[j]} = 1\right)}\right).$$

Proof. In an unconditional setting the weights are $1/n$ for $i = 1, \ldots, n$. Thus, $\widehat{S}^c_n(t)$ becomes

$$\widehat{S}^c_n(t) = \prod_{i=1}^n \left(1 - \frac{\delta_{[i]}\frac{1}{n}\mathbf{1}\left(T_{(i)} \leq t\right)}{\frac{1}{n}(n - i + 1) + \frac{1}{n}\sum_{j=1}^{i-1}\mathbf{1}\left(\xi_{[j]}\nu_{[j]} = 1\right)}\right).$$

□

Proof of Proposition 2.1

Proposition 2.1 The estimator $\widehat{S}^c_h(t \mid x)$ has the following properties.

1. If there is no known cure status, $\widehat{S}^c_h(t \mid x)$ reduces to $\widehat{S}_h(t \mid x)$.

   Proof. It is straightforward since $\xi_i\nu_i = 0$, $i = 1, \ldots, n$. □

2. The proposed estimator $\widehat{S}^c_h(t \mid x)$ reduces to $\widehat{S}_h(t \mid x)$ when computed with the *usual* observed times.

Proof. Let $\{T_i, i = 1, \ldots, n\}$ be the *actual* observed times, and define the *usual* observed times by means of the usual definition in survival analysis, that is, as $\widetilde{T}_i = \min(Y_i, C_i)$. Note that $\widetilde{T}_i = T_i$ if $\xi_i \nu_i = 0$, and $\widetilde{T}_i = \infty$ if the individual is known to be cured ($\xi_i \nu_i = 1$). It is straightforward to see that the $n = n_1 + n_2$ observations can be ordered so the first n_1 observations correspond to individuals with finite *usual* observed times $\widetilde{T}_i < \infty$ ($\xi_i \nu_i = 0$) and the remaining n_2 observations to individuals with *usual* observed time $\widetilde{T}_i = \infty$ ($\xi_i \nu_i = 1$). Therefore,

$$\begin{aligned}\widehat{S}_h^c\left(t \mid x\right) &= \prod_{i=1}^{n}\left(1 - \frac{\delta_{[i]} B_{h[i]}\left(x\right)\ \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n_1} B_{h[j]}\left(x\right) + \sum_{j=n_1+1}^{n} B_{h[j]}\left(x\right)}\right)\\ &= \prod_{i=1}^{n}\left(1 - \frac{\delta_{[i]} B_{h[i]}\left(x\right) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}\left(x\right)}\right) = \widehat{S}_h(t \mid x).\end{aligned}$$

This completes the proof. □

3. In the specific case when some individuals are observed as cured when their observed time exceeds a known fixed cure threshold, $\widehat{S}_h^c(t \mid x)$ reduces to $\widehat{S}_h(t \mid x)$.

Proof. Assume there exists a common specific known cure threshold $d_i = d$ for $i = 1, \ldots, n$. This implies that in the ordered sample, $\left\{\left(X_{[i]}, T_{(i)}, \delta_{[i]}, \xi_{[i]}\nu_{[i]}\right), i = 1, \ldots, n\right\}$, the n_1 first observations correspond to individuals with $T_{(i)} < d$ either not cured or with unknown cure status ($\xi_{[i]}\nu_{[i]} = 0$), and the remaining n_2 observations are cured individuals with $T_{(i)} \geq d$ and $\xi_{[i]}\nu_{[i]} = 1$. Therefore,

$$\begin{aligned}\widehat{S}_h^c\left(t \mid x\right) &= \prod_{i=1}^{n}\left(1 - \frac{\delta_{[i]} B_{h[i]}\left(x\right)\ \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n_1} B_{h[j]}\left(x\right) + \sum_{j=n_1+1}^{n} B_{h[j]}\left(x\right)}\right)\\ &= \prod_{i=1}^{n}\left(1 - \frac{\delta_{[i]} B_{h[i]}\left(x\right) \mathbf{1}\left(T_{(i)} \leq t\right)}{\sum_{j=i}^{n} B_{h[j]}\left(x\right)}\right) = \widehat{S}_h(t \mid x).\end{aligned}$$

This completes the proof. □

4. When there is no censoring, the estimator $\widehat{S}_h^c(t|x)$ reduces to the kernel type estimator of the conditional survival function.

Proof. Without censoring, $T_i = Y_i, \delta_i = 1$ and the cure status is always observed $\xi_i = 1$. In this situation, the $n = n_1 + m$ observations can be ordered and split into the n_1 uncured individuals with finite lifetimes Y_i, and the m cured individuals with lifetime $Y_i = \infty$. Thus,

$$\begin{aligned}
\widehat{S}_h^c(t \mid x) &= \prod_{i=1}^{n} \left(1 - \frac{B_{h[i]}\,(x)\,\mathbf{1}\left(Y_{(i)} \le t\right)}{\sum_{j=i}^{n} B_{h[j]}\,(x) + \sum_{j=1}^{i-1} B_{h[j]}\,(x)\,\mathbf{1}\left(\nu_{[j]} = 1\right)} \right) \\
&= \prod_{i=1}^{n} \left(1 - \frac{B_{h[i]}\,(x)\,\mathbf{1}\left(Y_{(i)} \le t\right)}{\sum_{j=i}^{n_1} B_{h[j]}\,(x) + \sum_{j=n_1+1}^{n} B_{h[j]}\,(x)} \right) \\
&= \prod_{i:Y_{(i)} \le t} \left(\frac{\sum_{j=i+1}^{n} B_{h[j]}\,(x)}{\sum_{j=i}^{n} B_{h[j]}\,(x)} \right).
\end{aligned}$$

Note that the kernel estimator of the survival function

$$\widetilde{S}_h(t \mid x) = \sum_{i=1}^{n} B_{h[i]}\,(x)\,\mathbf{1}(Y_{(i)} > t)$$

is a step function with jumps $B_{hi}\,(x)$ at the observations, Y_i. By defining $k = \max\{i : Y_{(i)} \le t\}$ i.e., $Y_{(k)} \le t$ and $Y_{(k+1)} > t$, one can write

$$\begin{aligned}
\prod_{i:Y_{(i)} \le t} \left(\frac{\sum_{j=i+1}^{n} B_{h[j]}\,(x)}{\sum_{j=i}^{n} B_{h[j]}\,(x)} \right) &= \prod_{i:Y_{(i)} \le t} \left(\frac{\widetilde{S}_h(Y_{(i)} \mid x)}{\widetilde{S}_h(Y_{(i-1)} \mid x)} \right) \\
&= \frac{\widetilde{S}_h(Y_{(1)} \mid x)}{1}\,\frac{\widetilde{S}_h(Y_{(2)}|x)}{\widetilde{S}_h(Y_{(1)} \mid x)} \cdots \frac{\widetilde{S}_h(Y_{(k)} \mid x)}{\widetilde{S}_h(Y_{(k-1)} \mid x)} \\
&= \widetilde{S}_h(Y_{(k)} \mid x) = \sum_{i=1}^{n} B_{h[i]}\,(x)\,\mathbf{1}(Y_{(i)} > t).
\end{aligned}$$

This completes the proof. □

5. In an unconditional setting and in the particular case where an individual is known to be cured only if the observed time is greater than a known fixed time, say d, $\widehat{S}_n^c\,(t)$ in (2.18) reduces to the generalized maximum likelihood estimator in Laska and Meisner (1992).

Proof. In the particular case where an individual is known to be cured only if the observed time is greater than a known fixed time, say d, with $n = n_1 + m$ observations, when m are identified as cured, the ordered observed lifetimes are $T_{(1)} \le \ldots \le T_{(n_1)}$ strictly lower than d, and the m cured individuals with $T_{(i)} \ge d$. Thus, $\widehat{S}_n^c(t)$ reduces to the generalized maximum likelihood estimator in Laska and Meisner (1992):

$$\widehat{S}_n^c(t) = \prod_{i=1}^n \left(1 - \frac{\delta_{[i]}\frac{1}{n}\mathbf{1}\left(T_{(i)} \le t\right)}{\frac{1}{n}(n_1 - i + 1) + \frac{1}{n}m}\right) = \prod_{i=1}^n \left(1 - \frac{\delta_{[i]}\mathbf{1}(T_{(i)} \le t)}{n - i + 1}\right).$$

This completes the proof. □

Proof of Proposition 2.2

Proposition 2.2 The estimators $1 - \widehat{F}_h^c(t \mid x)$ and $1 - \widehat{G}_h^c(t \mid x)$ in (2.16) and (2.17) verify that

$$1 - \widehat{H}_h^0(t \mid x) = \left(1 - \widehat{F}_h^c(t \mid x)\right)\left(1 - \widehat{G}_h^c(t \mid x)\right)$$

where

$$1 - \widehat{H}_h^0(t \mid x) = \sum_{i=1}^n B_{hi}(x)\,\mathbf{1}\left(T_i > t, \xi_i\nu_i = 0\right).$$

Proof. Consider the expression of the estimators $1-\widehat{F}_h^c(t \mid x)$ and $1-\widehat{G}_h^c(t \mid x)$ in (2.19) and (2.20) respectively, that is, computed with the *usual* times $\widetilde{T}_i = T_i$ if $\xi_i\nu_i = 0$ and $\widetilde{T}_i = \infty$ if $\xi_i\nu_i = 1$. Then, the product can be worked out as follows:

$$\begin{aligned}
&\left(1 - \widehat{F}_h^c(t \mid x)\right)\left(1 - \widehat{G}_h^c(t \mid x)\right)\\
&= \prod_{i=1}^n \left(1 - \frac{B_{h[i]}(x)\,\mathbf{1}\left(\widetilde{T}_{(i)} \le t\right)}{\sum_{j=i}^n B_{h[j]}(x)}\right)^{\delta_{[i]}} \left(1 - \frac{B_{h[i]}(x)\,\mathbf{1}\left(\widetilde{T}_{(i)} \le t\right)}{\sum_{j=i}^n B_{h[j]}(x)}\right)^{\left(1-\delta_{[i]}\right)}\\
&= \prod_{i=1}^n \left(1 - \frac{B_{h[i]}(x)\,\mathbf{1}\left(\widetilde{T}_{(i)} \le t\right)}{\sum_{j=i}^n B_{h[j]}(x)}\right)\\
&= \prod_{\substack{i=1,\\ \widetilde{T}_{(i)} \le t}}^n \left(1 - \frac{B_{h[i]}(x)}{\sum_{j=i}^n B_{h[j]}(x)}\right) = \prod_{\substack{i=1,\\ \widetilde{T}_{(i)} \le t}}^n \frac{\sum_{j=i+1}^n B_{h[j]}(x)}{\sum_{j=i}^n B_{h[j]}(x)}.
\end{aligned}$$

Let $k = \max\{i : \widetilde{T}_{(i)} \le t\}$, then

$$\begin{aligned}
&\left(1 - \widehat{F}_h^c(t \mid x)\right)\left(1 - \widehat{G}_h^c(t \mid x)\right)\\
&= \prod_{i=1}^k \frac{\sum_{j=i+1}^n B_{h[j]}(x)}{\sum_{j=i}^n B_{h[j]}(x)}\\
&= \frac{\sum_{j=2}^n B_{h[j]}(x)}{\sum_{j=1}^n B_{h[j]}(x)}\frac{\sum_{j=3}^n B_{h[j]}(x)}{\sum_{j=2}^n B_{h[j]}(x)} \cdots \frac{\sum_{j=k}^n B_{h[j]}(x)}{\sum_{j=k-1}^n B_{h[j]}(x)}\frac{\sum_{j=k+1}^n B_{h[j]}(x)}{\sum_{j=k}^n B_{h[j]}(x)}
\end{aligned}$$

$$= \frac{\sum_{j=k+1}^{n} B_{h[j]}(x)}{\sum_{j=1}^{n} B_{h[j]}(x)} = \sum_{j=k+1}^{n} B_{h[j]}(x) = \sum_{i=1}^{n} B_{hi}(x) \mathbf{1}\left(\widetilde{T}_i \leq t\right)$$
$$= \sum_{i=1}^{n} B_{hi}(x) \mathbf{1}(T_i \leq t, \xi_i \nu_i = 0) = 1 - \widehat{H}_h^0(t \mid x).$$

This completes the proof. □

Proof of Propositions 2.3

Proposition 2.3 The $1 - \widehat{F}_h^c(t \mid x)$ estimator in (2.16) is the nonparametric local maximum likelihood estimator of $1 - F(t \mid x)$.

Proof. The proof follows the argument in Theorem 2 in López-Cheda et al. (2017a) and Theorem 1 in Laska and Meisner (1992). To derive the expression of the local likelihood of the MCM, we consider the three potential cases for the ith observation:

Case 1: ($\delta_i = 1$). The event is observed and the individual is not cured. We observe $Y_i = t_i, \nu_i = 0$, with probability:

$$\begin{aligned}
&P(Y_i = t_i, C_i > t_i, \nu_i = 0 \mid X = x) \\
&= P(C_i > t_i \mid Y_i = t_i, \nu_i = 0, X = x) \\
&\quad \times P(Y_i = t_i \mid \nu_i = 0, X = x) P(\nu_i = 0 \mid X = x) \\
&= S_{C|Y,X,\nu=0}(t_i \mid x) \left(S_0(t_i^- \mid x) - S_0(t_i \mid x)\right) p(x),
\end{aligned}$$

where $S_{C|Y,X,\nu=0}(t \mid x)$ is the conditional survival function of the censoring variable C for the uncured individuals.

Case 2: ($\delta_i = 0, \xi_i\nu_i = 0$). The individual is censored and the cure status is unknown. We observe $C_i = t_i$, and ν_i is unknown, with probability:

$$\begin{aligned}
&P(Y_i > t_i, C_i = t_i \mid X = x) \\
&= P(Y_i > t_i, C_i = t_i \mid \nu_i = 1, X = x) P(\nu_i = 1 \mid X = x) \\
&\quad + P(C_i = t_i \mid Y_i > t_i, \nu_i = 0, X = x) \\
&\quad \times P(Y_i > t_i \mid \nu_i = 0, X = x) P(\nu_i = 0 \mid X = x) \\
&= f_{C|X,\nu=1}(t_i \mid x)(1 - p(x)) + f_{C|Y,X,\nu=0}(t_i \mid x) S_0(t_i \mid x) p(x),
\end{aligned}$$

where $f_{C|X,\nu=1}(t \mid x)$ and $f_{C|Y,X,\nu=0}(t \mid x)$ are the conditional density functions for the random variable C of the cured and uncured individuals, respectively.

Case 3: $(\delta_i = 0, \xi_i\nu_i = 1)$**.** The individual is censored and known to be cured. We observe $C_i = t_i, \nu_i = 1$, with probability

$$\begin{aligned}
&P\left(Y_i > t_i, C_i = t_i, \nu_i = 1 \mid X = x\right)\\
&= P\left(C_i = t_i \mid \nu_i = 1, X = x\right) P\left(\nu_i = 1 \mid X = x\right)\\
&= f_{C|X,\nu=1}\left(t_i \mid x\right)\left(1 - p\left(x\right)\right).
\end{aligned}$$

In the absence of specification of the distribution of X, the terms in the likelihood are weighted with the kernel weights $B_{h[i]}(x)$. Then, the local likelihood of the data is

$$\begin{aligned}
&L\left(X, T, \delta, \xi, \nu\right)\\
&= \prod_{i=1}^{n}\left[S_{C|Y,X,\nu=0}\left(T_{(i)} \mid x\right)\left(S_0(T_{(i)}^{-} \mid x) - S_0(T_{(i)} \mid x)\right)p\left(x\right)\right]^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)}\\
&\times\Bigg[f_{C|X,\nu=1}\left(T_{(i)} \mid x\right)\left(1 - p\left(x\right)\right)\\
&+ f_{C|Y,X,\nu=0}\left(T_{(i)} \mid x\right)S_0\left(T_{(i)} \mid x\right)p\left(x\right)\Bigg]^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=0\right)}\\
&\times\left[f_{C|X,\nu=1}\left(T_{(i)} \mid x\right)\left(1 - p\left(x\right)\right)\right]^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=1\right)}.
\end{aligned}$$

If the distribution of the random variable C is conditionally independent of Y and the cure status ν given the covariate $X = x$, then

$$\begin{aligned}
L\left(X, T, \delta, \xi, \nu\right) = \prod_{i=1}^{n}&\left[q_i(x)p\left(x\right)\right]^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)}\\
&\times\left(1 - p\left(x\right) + S_0\left(T_{(i)} \mid x\right)p\left(x\right)\right)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=0\right)}\\
&\times\left(1 - p\left(x\right)\right)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=1\right)}\\
&\times\left(1 - \sum_{j=1}^{i-1} g_j\left(x\right)\right)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)} g_i\left(x\right)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0\right)},
\end{aligned}\tag{A.1}$$

where, for $i = 1, \ldots, n$, $q_i\left(x\right) = S_0(T_{(i)}^{-} \mid x) - S_0(T_{(i)} \mid x)$ are the increments of $S_0\left(t \mid x\right)$, and $g_i\left(x\right) = G(T_{(i)} \mid x) - G(T_{(i)}^{-} \mid x)$ the increments of $G\left(t \mid x\right)$. Let $P_i\left(x\right) = p\left(x\right)q_i\left(x\right)$ be the increments of $S\left(t \mid x\right)$, then $\sum_{i=1}^{n} P_i\left(x\right) = p\left(x\right)$. Maximizing (A.1) is equivalent to maximizing the likelihood

$$L\left(X, T, \delta, \xi, \nu\right) = \prod_{i=1}^{n} P_i\left(x\right)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)}\left(1 - \sum_{j=1}^{i-1} P_j\left(x\right)\right)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=0\right)}$$

$$\times\left(1-\sum_{j=1}^{n}P_{j}\left(x\right)\right)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=1\right)}. \tag{A.2}$$

Further, consider the functions $\lambda_{i}\left(x\right)=P_{i}\left(x\right)/\left(1-\sum_{j=1}^{i-1}P_{j}\left(x\right)\right)$ satisfying

$$1-\sum_{j=1}^{k}P_{j}\left(x\right)=\prod_{j=1}^{k}(1-\lambda_{j}\left(x\right)). \tag{A.3}$$

Then, the increments $P_{i}\left(x\right)$ can be written in terms of $\lambda_{i}\left(x\right)$:

$$P_{i}\left(x\right)=\lambda_{i}\left(x\right)\left(1-\sum_{j=1}^{i-1}P_{j}\left(x\right)\right)=\lambda_{i}\left(x\right)\prod_{j=1}^{i-1}\left(1-\lambda_{j}\left(x\right)\right). \tag{A.4}$$

By substituting (A.3) and (A.4) in (A.2), the likelihood (A.2) is

$$\begin{aligned}L\left(X,T,\delta,\xi,\nu;p,S_{0}\right)=&\prod_{i=1}^{n}\lambda_{i}\left(x\right)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)}\prod_{i=1}^{n}\left[\prod_{j=1}^{i-1}\left(1-\lambda_{j}\left(x\right)\right)\right]^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)}\\&\times\prod_{i=1}^{n}\left[\prod_{j=1}^{i-1}\left(1-\lambda_{j}\left(x\right)\right)\right]^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=0\right)}\\&\times\prod_{i=1}^{n}\left[\prod_{j=1}^{n}\left(1-\lambda_{j}\left(x\right)\right)\right]^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=0,\xi_{[i]}\nu_{[i]}=1\right)}.\end{aligned}$$

Taking into account that $\prod_{i=1}^{n}\left[\prod_{j=1}^{i-1}a_{j}\right]^{b_{i}}=\prod_{i=1}^{n}a_{i}^{\sum_{j=i+1}^{n}b_{j}}$, where a_{i} and b_{i}, $i=1,\dots,n$, are arbitrary sequences of nonnegative numbers, the likelihood becomes

$$\begin{aligned}&L\left(X,T,\delta,\xi,\nu;p,S_{0}\right)\\&=\prod_{i=1}^{n}\lambda_{i}\left(x\right)^{B_{h[i]}(x)\mathbf{1}\left(\delta_{[i]}=1\right)}\\&\times\prod_{i=1}^{n}\left(1-\lambda_{i}\left(x\right)\right)^{\sum_{j=i+1}^{n}B_{h[j]}(x)\mathbf{1}\left(\xi_{[j]}\nu_{[j]}=0\right)+\sum_{j=1}^{n}B_{h[j]}(x)\mathbf{1}\left(\delta_{[j]}=0,\xi_{[j]}\nu_{[j]}=1\right)}.\end{aligned}$$

Maximizing the likelihood $L\left(X,T,\delta,\xi,\nu;p,S_{0}\right)$ is equivalent to maximizing the local log-likelihood:

$$\begin{aligned}\Psi(\lambda_{1}\left(x\right),\dots,\lambda_{n}\left(x\right))=\sum_{i=1}^{n}\Bigg[&B_{h[i]}\left(x\right)\mathbf{1}\left(\delta_{[i]}=1\right)\log\lambda_{i}\left(x\right)\\&+\Bigg(\sum_{j=i+1}^{n}B_{h[j]}\left(x\right)\mathbf{1}\left(\xi_{[j]}\nu_{[j]}=0\right)\end{aligned}$$

$$+\sum_{j=1}^{n} B_{h[j]}\left(x\right)\mathbf{1}\left(\delta_{[j]}=0,\xi_{[j]}\nu_{[j]}=1\right)\Big)\log\left(1-\lambda_i\right)\Big]$$

subject to

$$\prod_{i=1}^{n}(1-\lambda_i\left(x\right))=1-p\left(x\right). \tag{A.5}$$

The maximizer $\lambda_i\left(x\right)$ of the log-likelihood is

$$\widehat{\lambda}_i\left(x\right)=\frac{B_{h[i]}\left(x\right)\mathbf{1}\left(\delta_{[i]}=1\right)}{\sum_{j=i}^{n}B_{h[j]}\left(x\right)+\sum_{j=1}^{i-1}B_{h[j]}\left(x\right)\mathbf{1}\left(\xi_{[j]}\nu_{[j]}=1\right)}.$$

In virtue of (A.4), the estimator $\widehat{S}_h^c(t\mid x)$, computed by forming the product of $\widehat{\lambda}_i$'s such that $T_{(i)}\leq t$, is the nonparametric maximum likelihood estimator of $S(t\mid x)$. On the hand, the estimator $1-\widehat{p}_h^c(x)$, obtained by replacing $\widehat{\lambda}_i$'s in (A.5), is the nonparametric maximum likelihood estimator of $1-p(x)$. □

Proof of Theorem 2.1

Theorem 2.1 Suppose that Assumptions $1-9$ are satisfied. Then, for $x\in I, t\in[a,b]$ one has

$$\widehat{\Lambda}_h^c\left(t\mid x\right)-\Lambda\left(t\mid x\right)=\sum_{i=1}^{n}\widetilde{B}_{hi}\left(x\right)\zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right)+R_{n1}\left(t,x\right),$$

with

$$\begin{aligned}\zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right)&=\frac{\mathbf{1}\left(T_i\leq t,\delta_i=1\right)}{J(T_i^-\mid x)}\\&\quad-\int_0^t\left(\mathbf{1}\left(T_i\geq v\right)+\mathbf{1}\left(T_i<v,\xi_i\nu_i=1\right)\right)\frac{dH^1\left(v\mid x\right)}{J^2\left(v^-\mid x\right)},\\\widetilde{B}_{hi}\left(x\right)&=\frac{1}{m\left(x\right)}\frac{1}{nh}K\left(\frac{x-X_i}{h}\right),\end{aligned}$$

where $R_{n1}\left(t,x\right)$ satisfies

$$\sup_{a\leq t\leq b,x\in I}\mid R_{n1}\left(t,x\right)\mid=O\left(\left(nh\right)^{-3/4}\left(\log n\right)^{3/4}\right)\quad\text{a.s.}$$

Proof. The difference $\widehat{\Lambda}_h^c(t \mid x) - \Lambda(t \mid x)$ can be decomposed as follows:

$$\begin{aligned}
&\int_0^t \frac{d\widehat{H}_h^1(v \mid x)}{\widehat{J}_h(v^- \mid x)} - \int_0^t \frac{dH^1(v \mid x)}{J(v^- \mid x)} \\
=& \int_0^t \frac{d\widehat{H}_h^1(v \mid x)}{J(v^- \mid x)} - \int_0^t \frac{\widehat{J}_h(v^- \mid x)}{J^2(v^- \mid x)} dH^1(v \mid x) \\
&+ \int_0^t \frac{\left(J(v^- \mid x) - \widehat{J}_h(v^- \mid x)\right)^2}{\widehat{J}_h(v^- \mid x) J^2(v^- \mid x)} dH^1(v \mid x) \\
&+ \int_0^t \left(\frac{1}{\widehat{J}_h(v^- \mid x)} - \frac{1}{J(v^- \mid x)}\right)\left(d\widehat{H}_h^1(v \mid x) - dH^1(v \mid x)\right). \qquad \text{(A.6)}
\end{aligned}$$

The first term in (A.6) is

$$\int_0^t \frac{d\widehat{H}_h^1(v \mid x)}{J(v^- \mid x)} = \sum_{i=1}^n B_{hi}(x) \frac{\mathbf{1}(T_i \le t, \delta_i = 1)}{J(T_i^- \mid x)}. \qquad \text{(A.7)}$$

The second term in (A.6) can be written as

$$\begin{aligned}
&\int_0^t \frac{\widehat{J}_h(v^- \mid x)}{J^2(v^- \mid x)} dH^1(v \mid x) \\
=& \sum_{i=1}^n B_{hi}(x) \int_0^t \left(\mathbf{1}(T_i \ge v) + \mathbf{1}(T_i < v, \xi_i \nu_i = 1)\right) \frac{dH^1(v \mid x)}{J^2(v^- \mid x)}. \qquad \text{(A.8)}
\end{aligned}$$

The order of the third and fourth terms in (A.6) are studied, denoting them by $R_1(t,x)$ and $R_2(t,x)$, respectively. Theorem 1 in Iglesias-Pérez and González-Manteiga (1999) is applied, under Assumptions $1-3$, and $8-9$, which also holds for $J(t \mid x)$ such that $\sup_{a \le t \le b, x \in I} \mid \widehat{J}_h(t \mid x) - J(t \mid x) \mid = O((nh)^{-1/2}(\log n)^{1/2})$. Note that Assumption 1. (ii) ensures the existence of a constant $\theta > 0$ such that $J(t \mid x) \ge \theta > 0$ for (t,x) in $[a,b] \times I_\varepsilon$. As n becomes sufficiently large, the remainder term $R_1(t,x)$ is bounded by

$$\begin{aligned}
\sup_{a \le t \le b, x \in I} \mid R_1(t,x) \mid &\le \sup_{a \le t \le b, x \in I} \int_0^t \left| \frac{\left(J(v \mid x) - \widehat{J}_h(v \mid x)\right)^2}{\widehat{J}_h(v \mid x) J^2(v \mid x)} dH^1(v \mid x) \right| \\
&\le \left(\sup_{a \le t \le b, x \in I} \mid J(t \mid x) - \widehat{J}_h(t \mid x) \mid \right)^2 \sup_{a \le t \le b, x \in I} \int_0^t \frac{1}{\theta^3} dH^1(v \mid x) \\
&\le \frac{1}{\theta^3} \left(\sup_{a \le t \le b, x \in I} \mid J(t \mid x) - \widehat{J}_h(t \mid x) \mid \right)^2.
\end{aligned}$$

Following Lemma 5 in Iglesias-Pérez and González-Manteiga (1999), under Assumptions $1-3$ and $8-9$, which not only holds for the conditional survival functions like $1 - H(t \mid x)$ but also for the conditional subdistribution functions $H^1(t \mid x)$ and

$H^{11}(t \mid x)$ (see Remark 2 in Iglesias-Pérez and González-Manteiga (1999) and the proof of Theorem 2.1 in Dabrowska (1989)), it is shown that

$$\sup_{a\leq t\leq b, x\in I} \mid R_1(t,x) \mid = O((nh)^{-1}\log n) \text{ a.s.} \tag{A.9}$$

Now let us study the remainder term $R_2(t,x)$, which is bounded by

$$\begin{aligned}
&\sup_{a\leq t\leq b, x\in I} \mid R_2(t,x) \mid \\
&\leq \sup_{a\leq t\leq b, x\in I}\left| \int_0^t \left(\frac{1}{\widehat{J}_h\left(v \mid x\right)} - \frac{1}{J\left(v \mid x\right)} \right) d\widehat{H}_h^1\left(v \mid x\right) \right| \\
&\quad + \sup_{a\leq t\leq b, x\in I}\left| \int_0^t \left(\frac{1}{\widehat{J}_h\left(v \mid x\right)} - \frac{1}{J\left(v \mid x\right)} \right) dH^1\left(v \mid x\right) \right| \\
&\leq \sup_{a\leq t\leq b, x\in I} \int_0^t \left| \left(\frac{1}{\widehat{J}_h\left(v \mid x\right)} - \frac{1}{J\left(v \mid x\right)} \right) \left(d\widehat{H}_h^1\left(v \mid x\right) - dH^1\left(v \mid x\right)\right) \right| \\
&\quad + 2 \sup_{a\leq t\leq b, x\in I} \int_0^t \left| \left(\frac{1}{\widehat{J}_h\left(v \mid x\right)} - \frac{1}{J\left(v \mid x\right)} \right) \right| dH^1\left(v \mid x\right) \\
&\leq \frac{k_n}{\theta^2} \sup_{a\leq t\leq b, x\in I} \left(\widehat{J}_h\left(t \mid x\right) - J\left(t \mid x\right)\right) \max_{1\leq i\leq k_n} \left| \left(\widehat{H}_h^1\left(t_{i+1} \mid x\right) - H^1\left(t_{i+1} \mid x\right)\right) \right. \\
&\quad \left. - \left(\widehat{H}_h^1\left(t_i \mid x\right) - H^1\left(t_i \mid x\right)\right) \right| \\
&\quad + 2k_n \max_{1\leq i\leq k_n} \sup_{t_i\leq t\leq t_{i+1}} \left| \left(\frac{1}{\widehat{J}_h\left(t \mid x\right)} - \frac{1}{J\left(t \mid x\right)} \right) - \left(\frac{1}{\widehat{J}_h\left(t_i \mid x\right)} - \frac{1}{J\left(t_i \mid x\right)} \right) \right|,
\end{aligned}$$

where $([t_i, t_{i+1}])_{i=1}^{k_n}$ denotes a partition of the interval $[a,b]$ in k_n intervals with $a = t_1 < \ldots < t_{k_n+1} = b$ and $k_n = O\left((nh)^{-1}\log n\right)^{-1/2}$. In order to show that $R_2(t,x)$ is negligible, the arguments similar to part (c) in Theorem 2 of Iglesias-Pérez and González-Manteiga (1999) are considered under Assumptions $1-9$. Thus, it can be proved that

$$\sup_{a\leq t\leq b, x\in I} |R_2(t,x)| = O\left((nh)^{-3/4}(\log n)^{3/4}\right) \text{ a.s.} \tag{A.10}$$

Finally, collecting (A.7), (A.8), (A.9) and (A.10), then (A.6) can be written as follows:

$$\widehat{\Lambda}_h^c\left(t \mid x\right) - \Lambda\left(t \mid x\right) = \sum_{i=1}^n B_{hi}\left(x\right) \zeta\left(T_i, \delta_i, \xi_i, \nu_i, t, x\right) + \widetilde{R}_{n1}\left(t, x\right), \tag{A.11}$$

where $\zeta\left(T_i, \delta_i, \xi_i, \nu_i, t, x\right)$ was defined in (2.21) and $\widetilde{R}_{n1}(t,x) = R_1(t,x) + R_2(t,x)$ verifies

$$\sup_{a\leq t\leq b, x\in I} \mid \widetilde{R}_{n1}\left(t,x\right) \mid = O\left((nh)^{-3/4}\left(\log n\right)^{3/4}\right) \text{ a.s.}$$

The sum in (A.11) can be decomposed into three terms:

$$\widehat{\Lambda}_h^c\left(t \mid x\right) - \Lambda\left(t \mid x\right) = \sum_{i=1}^n \widetilde{B}_{hi}\left(x\right)\zeta\left(T_i, \delta_i, \xi_i, \nu_i, t, x\right) + \widetilde{R}_{n1}(t,x) + R_3(t,x),$$

where

$$R_3(t,x) = \sum_{i=1}^n \frac{m(x) - \widehat{m}_h(x)}{m(x)\widehat{m}_h(x)} \frac{1}{nh} K\left(\frac{x - X_i}{h}\right) \zeta\left(T_i, \delta_i, \xi_i, \nu_i, t, x\right),$$

with $\widehat{m}_h(x)$ the kernel estimator of the density function of X, $m(x)$, and $\widetilde{B}_{hi}\left(x\right) = K((x - X_i)/h)/(m(x)nh)$. The next result can be proved by following similar arguments as that used in the proof of Lemma 5 in Iglesias-Pérez and González-Manteiga (1999), under Assumptions $1-3$ and $8-9$, one obtain

$$\sup_{x\in I} \mid m(x) - \widehat{m}_h(x) \mid = O\left(h^2 + (nh)^{-1/2}\left(\log\log n\right)^{1/2}\right) \text{ a.s.} \tag{A.12}$$

Moreover, after applying similar arguments as in Theorem 2.1 of Dabrowska (1989) under the assumptions of this theorem, it can be shown that

$$\begin{aligned} &\sup_{a\leq t\leq b, x\in I} \left| \frac{1}{\widehat{m}_h(x)} \sum_{i=1}^n \frac{1}{nh} K\left(\frac{x - X_i}{h}\right) \zeta\left(T_i, \delta_i, \xi_i, \nu_i, t, x\right) \right| \\ &= O\left(h^2 + (nh)^{-1/2}\left(\log\log n\right)^{1/2}\right) \text{ a.s.} \end{aligned} \tag{A.13}$$

Thus, from (A.12) and (A.13), under Assumption 1 and using the condition $nh \to \infty$, it is concluded that

$$\sup_{a\leq t\leq b, x\in I} \mid R_3(t,x) \mid = O\left(h^4 + (nh)^{-1}\log\log n\right) \text{ a.s.}$$

□

Proof of Theorem 2.2

Theorem 2.2 Suppose that Assumptions 1 − 9 hold. Then, for $x \in I$ and $t \in [a, b]$,

$$\widehat{F}_h^c (t \mid x) - F (t \mid x) = (1 - F (t \mid x)) \sum_{i=1}^{n} \widetilde{B}_{hi} (x) \zeta (T_i, \delta_i, \xi_i, \nu_i, t, x) + R_{n2} (t, x)$$

where $\zeta (T_i, \delta_i, \xi_i, \nu_i, t, x)$ is defined in (2.21), $\widetilde{B}_{hi}(x)$ in (2.22) and $R_{n2} (t, x)$ satisfies

$$\sup_{a \leq t \leq b, x \in I} \mid R_{n2} (t, x) \mid = O \left((nh)^{-3/4} (\log n)^{3/4} \right) \text{ a.s.}$$

Proof. Start by writing

$$\widehat{F}_h^c (t \mid x) - F (t \mid x) = \widetilde{R}_1 (t, x) + \left[1 - \exp \left(-\widehat{\Lambda}_h^c (t \mid x) \right) \right] - F (t \mid x) ,$$

where $\widetilde{R}_1 (t, x) = \widehat{F}_h^c (t \mid x) - \left[1 - \exp \left(-\widehat{\Lambda}_h^c (t \mid x) \right) \right]$. By a Taylor's expansion of the exponential function around $-\Lambda(t \mid x)$, we have

$$\widehat{F}_h^c (t \mid x) - F (t \mid x) = (1 - F (t \mid x)) \left(\widehat{\Lambda}_h^c (t \mid x) - \Lambda (t \mid x) \right) + \widetilde{R}_1 (t, x) + \widetilde{R}_2 (t, x) ,$$

where

$$\widetilde{R}_2 (t, x) = -\frac{1}{2} \exp(-\Lambda^*(t \mid x)) \left(\widehat{\Lambda}_h^c (t \mid x) - \Lambda (t \mid x) \right)^2 .$$

Note that $\Lambda^*(t \mid x)$ is on the line segment between $\widehat{\Lambda}_h^c (t \mid x)$ and $\Lambda (t \mid x)$. Under Assumptions 1 − 9, and arguing similarly as in Theorem 2 (c) of Iglesias-Pérez and González-Manteiga (1999), it suffices to show

$$\sup_{a \leq t \leq b, x \in I} \mid \widetilde{R}_1 (t, x) \mid = O \left((nh)^{-1} \right) \text{ a.s.}$$

Making use now of the strong consistency results for $\widehat{\Lambda}_h^c (t \mid x)$ in Corollary 2.1, then

$$\sup_{a \leq t \leq b, x \in I} \mid \widetilde{R}_2 (t, x) \mid = O \left((nh)^{-1} \log n \right) \text{ a.s.}$$

The proof of Theorem 2.2 is concluded after applying Theorem 2.1. □

Proof of Corollary 2.1

Corollary 2.1 Suppose that Assumptions $1-9$ hold. Then, for $x \in I$ and $t \in [a,b]$,

$$\sup_{a\leq t\leq b, x\in I} \mid \widehat{\Lambda}_h^c\left(t \mid x\right) - \Lambda\left(t \mid x\right) \mid = O\left(\left(nh\right)^{-1/2}\left(\log n\right)^{1/2}\right) \quad \text{a.s.},$$

and

$$\sup_{a\leq t\leq b, x\in I} \mid \widehat{F}_h^c\left(t \mid x\right) - F\left(t \mid x\right) \mid = O\left(\left(nh\right)^{-1/2}\left(\log n\right)^{1/2}\right) \quad \text{a.s.}$$

Proof. The dominant part of $\widehat{\Lambda}_h^c(t \mid x) - \Lambda(t \mid x)$ in Theorem 2.1 verifies

$$\begin{aligned}
&\sum_{i=1}^n B_{hi}(x)\zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right)\\
=&\int_0^t \frac{d\widehat{H}_h^1\left(v \mid x\right)}{J\left(v^- \mid x\right)} - \int_0^t \frac{\widehat{J}_h\left(v^- \mid x\right)}{J^2\left(v^- \mid x\right)} dH^1\left(v \mid x\right)\\
=&\int_0^t \frac{d\widehat{H}_h^1\left(v \mid x\right) - dH^1(v \mid x)}{J\left(v^- \mid x\right)} - \int_0^t \frac{\widehat{J}_h\left(v^- \mid x\right) - J(v^- \mid x)}{J^2\left(v^- \mid x\right)} dH^1\left(v \mid x\right)\\
=&\left[\frac{\widehat{H}_h^1\left(v \mid x\right) - H^1\left(v \mid x\right)}{J\left(v^- \mid x\right)}\right]_0^t + \int_0^t \frac{\widehat{H}_h^1\left(v \mid x\right) - H^1\left(v \mid x\right)}{J^2\left(v^- \mid x\right)} dJ(v \mid x)\\
&- \int_0^t \frac{\widehat{J}_h\left(v^- \mid x\right) - J(v^- \mid x)}{J^2\left(v^- \mid x\right)} dH^1\left(v \mid x\right)\\
\leq&\frac{1}{\theta}\sup_{a\leq t\leq b, x\in I} \mid \widehat{H}_h^1\left(t \mid x\right) - H^1\left(t \mid x\right) \mid + \frac{1}{\theta}\sup_{a\leq t\leq b, x\in I} \mid \widehat{H}_h^1\left(t \mid x\right) - H^1\left(t \mid x\right) \mid\\
&- \frac{1}{\theta^2}\sup_{a\leq t\leq b, x\in I} \mid \widehat{J}_h\left(t \mid x\right) - J\left(t \mid x\right) \mid .
\end{aligned}$$

The last three terms in the inequality are bounded by applying Lemma 5 in Iglesias-Pérez and González-Manteiga (1999) under Assumptions $1-3$ and $8-9$, which holds not only for conditional survival functions like $1 - H(t \mid x)$, but also for conditional subdistribution functions as $H^1(t \mid x)$ and $H^{11}(t \mid x)$ (see Remark 2 in Iglesias-Pérez and González-Manteiga (1999) and the proof of Theorem 2.1 in Dabrowska (1989)). As a consequence, the dominant term of $\widehat{\Lambda}_h^c(t \mid x) - \Lambda(t \mid x)$ is bounded by

$$\sup_{a\leq t\leq b, x\in I} \mid \sum_{i=1}^n B_{hi}(x)\zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right) \mid = O\left(\left(nh\right)^{-1/2}\left(\log n\right)^{1/2}\right).$$

Using the results of Theorem 2.2 it is straightforward to prove the second part of this corollary. □

Proof of Proposition 2.4

Proposition 2.4 Suppose that Assumptions 1−9 hold. Then, the asymptotic bias and variance of the dominant term of $1-\widehat{F}_h^c\left(t \mid x\right)$ are, respectively,

$$\begin{aligned}
\mu_{h,c}(t,x) &= h^2 B_c(t,x) + O\left(h^4\right),\\
\sigma^2_{h,c}(t,x) &= (nh)^{-1} s^2_c(t,x) + O(n^{-1}h),
\end{aligned}$$

with

$$\begin{aligned}
B_c(t,x) &= \frac{(1-F\left(t \mid x\right))(2\Phi_c'\left(x,t,x\right) m'\left(x\right) + \Phi_c''\left(x,t,x\right) m\left(x\right))d_K}{2m\left(x\right)},\\
s^2_c(t,x) &= \frac{(1-F\left(t \mid x\right))^2 \Phi_1^c\left(x,t,x\right) c_K}{m\left(x\right)},
\end{aligned}$$

where $d_K = \int v^2 K(v)dv, c_K = \int K^2(v)dv$,

$$\begin{aligned}
\Phi_c\left(y,t,x\right) &= \mathrm{E}\left(\zeta\left(T,\delta,\xi,\nu,t,x\right) \mid X=y\right),\\
\Phi_1^c\left(y,t,x\right) &= \mathrm{E}\left(\zeta^2\left(T,\delta,\xi,\nu,t,x\right) \mid X=y\right),
\end{aligned}$$

with $\zeta\left(T,\delta,\xi,\nu,t,x\right)$ given in (2.21). Besides, $\Phi_c'\left(y,t,x\right)$ and $\Phi_c''\left(y,t,x\right)$ are the first and second derivatives of $\Phi_c\left(y,t,x\right)$ with respect to y.

Proof. From Theorem 2.2, the bias of the dominant term in the iid expression of $1-\widehat{F}_h^c\left(t \mid x\right)$ is asymptotically equal to the expected value of

$$\frac{(nh)^{-1}\left(1-F\left(t \mid x\right)\right)}{m\left(x\right)} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right) = I + II \tag{A.14}$$

where

$$\begin{aligned}
I = \frac{(nh)^{-1}\left(1-F\left(t \mid x\right)\right)}{m\left(x\right)} &\left[\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right)\right.\\
&\left. - \mathrm{E}\left(\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right)\right)\right], \qquad \text{(A.15)}\\
II = \frac{(nh)^{-1}(1-F\left(t \mid x\right))}{m\left(x\right)} &\mathrm{E}\left(\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right)\right). \qquad \text{(A.16)}
\end{aligned}$$

Since $E(I)=0$, the asymptotic bias is II. Using Lemmas D.1 and D.2,

$$II=\frac{h^2\left(1-F\left(t\mid x\right)\right)\left(\Phi_c''\left(x,t,x\right)m\left(x\right)+2\Phi_c'\left(x,t,x\right)m'\left(x\right)\right)d_K}{2m\left(x\right)}+O\left(h^4\right),$$

with $\Phi_c'(y,t,x)$ and $\Phi_c''(y,t,x)$ the first and second derivatives of $\Phi_c\left(y,t,x\right)$ with respect to y. Recalling (A.14), the asymptotic variance of the dominant term in the iid expression of $1-\widehat{F}_h^c(t\mid x)$ is

$$\operatorname{Var}\left(I\right)=\frac{\left(1-F\left(t\mid x\right)\right)^2}{m^2(x)}(V_1-V_2), \tag{A.17}$$

where

$$\begin{aligned}V_1&=\frac{1}{nh^2}\mathrm{E}\left(K^2\left(\frac{x-X}{h}\right)\zeta^2\left(T,\delta,\xi,\nu,t,x\right)\right),\\V_2&=\frac{1}{nh^2}\left[\mathrm{E}\left(K\left(\frac{x-X}{h}\right)\zeta\left(T,\delta,\xi,\nu,t,x\right)\right)\right]^2.\end{aligned}$$

From Lemmas D.1 and D.2, V_2 reduces to

$$V_2=\frac{1}{4}\frac{h^2}{n}d_K^2\left(\frac{\Phi_c''\left(x,t,x\right)m\left(x\right)+2\Phi_c'\left(x,t,x\right)m'\left(x\right)}{m\left(x\right)}\right)^2+O\left(\frac{h^4}{n}\right). \tag{A.18}$$

As for V_1, let us define $\Phi_1^c\left(y,t,x\right)=E\left(\zeta^2\left(T,\delta,\xi,\nu,t,x\right)\mid X=y\right)$. Then, after a change of variable and a Taylor's expansion (as in the proof of Lemma D.1) we obtain

$$V_1=\frac{1}{nh}\Phi_1^c\left(x,t,x\right)m\left(x\right)c_K+\frac{1}{2}\frac{h}{n}e_K\frac{d^2}{dy^2}\left(\Phi_1^c\left(y,t,x\right)m\left(y\right)\right)|_{y=x}+O\left(n^{-1}h^3\right) \tag{A.19}$$

where $e_K=\int v^2K^2(v)dv$. The proof concludes by substituting (A.18) and (A.19) into (A.17). □

Proof of Theorem 2.3.

Theorem 2.3 Suppose that Assumptions $1-9$ are satisfied. For $x\in I$ and $t\in[a,b]$, it follows that:

(i) If $nh^5\to 0$ and $(\log n)^3/nh\to 0$, then

$$(nh)^{1/2}\left(\widehat{F}_h^c(t\mid x)-F(t\mid x)\right)\xrightarrow{d}N(0,s_c^2(t,x)).$$

(ii) If $nh^5 \to C^5 > 0$, then

$$(nh)^{1/2}\left(\widehat{F}_h^c(t \mid x) - F(t \mid x)\right) \xrightarrow{d} N(C^{5/2}B_c(t,x), s_c^2(t,x)),$$

with $B_c(t,x)$ given in (2.25), $s_c^2(t,x)$ in (2.26) and C is a constant.

Proof. From Theorem 2.2, we consider

$$\begin{aligned}&(nh)^{1/2}\left(\widehat{F}_h^c\left(t \mid x\right) - F\left(t \mid x\right)\right)\\ &= (nh)^{1/2}\left(1 - F\left(t \mid x\right)\right)\sum_{i=1}^{n}\widetilde{B}_{hi}\left(x\right)\zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right) + (nh)^{1/2}R_{n2}\left(t,x\right)\end{aligned}$$

with $\zeta(T,\delta,\xi,\nu,t,x)$ and $R_{n2}(t,x)$ given in (2.21) and (2.23), respectively. The condition $(\log n)^3/nh \to 0$ implies that $(nh)^{1/2}(\log n/nh)^{3/4} \to 0$, so the remainder term $(nh)^{1/2}R_{n2}(t,x)$ is negligible. Consequently, the asymptotic distribution of $(nh)^{1/2}\left(\widehat{F}_h^c\left(t \mid x\right) - F\left(t \mid x\right)\right)$ is that of

$$(nh)^{1/2}\frac{1 - F\left(t \mid x\right)}{m\left(x\right)}\sum_{i=1}^{n}\frac{1}{nh}K\left(\frac{x - X_i}{h}\right)\zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right) = (nh)^{1/2}(I + II), \quad \text{(A.20)}$$

where I and II are given in (D.9) and (A.16). Under the assumption $nh^5 \to 0$, we have $(nh)^{1/2}II = o(1)$. Therefore, the asymptotic distribution of (A.20) is that of $(nh)^{1/2}I$. Let us define $(nh)^{1/2}I = \sum_{i=1}^{n}\eta_{i,h}(t,x)$, where

$$\begin{aligned}\eta_{i,h}(t,x) = \frac{(nh)^{-1/2}\left(1 - F\left(t \mid x\right)\right)}{m\left(x\right)}\Bigg[&K\left(\frac{x - X_i}{h}\right)\zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right)\\ &- \mathrm{E}\left(K\left(\frac{x - X_i}{h}\right)\zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right)\right)\Bigg],\end{aligned}$$

is a sequence of n independent random variables with mean 0. Note that

$$\begin{aligned}\mathrm{Var}(\eta_{i,h}(t,x)) = h\mathrm{Var}(I) &= \frac{1}{n}\frac{\left(1 - F(t \mid x)\right)^2}{m(x)}\Phi_1^c(x,t,x)c_K + O\left(\frac{h^2}{n}\right)\\ &= \frac{1}{n}s_c^2(t,x) + O\left(\frac{h^2}{n}\right)\end{aligned}$$

with $\mathrm{Var}(I)$ in (A.17) and $s_c^2(t,x)$ in (2.30). Since $\mathrm{Var}(\eta_{i,h}(t,x)) < \infty$ for $i = 1,\ldots,n$ and $\mathrm{Var}(\eta_h(t,x)) = \sum_{i=1}^{n}\mathrm{Var}(\eta_{i,h}(t,x))$ is positive, then we can apply Lindeberg's theorem (Billingsley, 1968) to obtain

$$\frac{\sum_{i=i}^{n} \eta_{i,h}(t,x)}{s_c^2(t,x)} \to N(0,1) \quad \text{in distribution.}$$

Therefore, $(nh)^{1/2}\left(\widehat{F}_h^c(t|x) - F(t \mid x)\right) \to N(0, s_c^2(t,x))$ in distribution. This proves (i). In parallel to the proof (i) we can prove (ii) as follows, note that if $nh^5 \to C^5$ then the bias term is $(nh)^{1/2}II = (nh)^{1/2}(h^2 B_c(t,x) + O(h^4)) = (nh^5)^{1/2} B_c(t,x) + O((nh^9)^{1/2})$ with $B_c(t,x)$ in (2.25). Thus,
$(nh)^{1/2}\left(\widehat{F}_h^c(t \mid x) - F(t \mid x)\right) \to N(C^{5/2}B_c(t,x), s_c^2(t,x))$ in distribution. □

Proof of Proposition 2.5

Proposition 2.5 Assume there are no ties in the observed times $\{T_1, \ldots, T_n\}$. Then, the simple weighted bootstrap and the obvious bootstrap are equivalent.

Proof. To prove the equivalence of the two resampling methods, we will follow that of Efron (1981), for right censored data without cured observations and without covariate. Specifically, we will prove that the conditional distribution of $(T_i^*, \delta_i^*, \xi_i^* \nu_i^*)$ defined by (2.35) – (2.37) given X_i^* is $\widehat{F}_g^c(. \mid X_i^*)$ in (2.33).

Let us consider that there are $m < n$ observations known to be cured. For the ease of convenience, the sample $\{(T_i, \delta_i, \xi_i \nu_i), i = 1, \ldots, n\}$ is ordered so the observations with unknown cure status come first, and it is separated into two subsamples: $\{(T_i, \delta_i, \xi_i \nu_i = 0), i = 1, \ldots, n-m\}$ and $\{(T_i, \delta_i = 0, \xi_i \nu_i = 1),$
$i = n-m+1, \ldots, n\}$

Let T_i^* be defined by (2.35), where $Y_i^* \sim \widehat{F}_g^c(t \mid X_i^*)$, $C_i^* \sim \widehat{G}_g^c(t \mid X_i^*)$ and $C_{0i}^* \sim \widehat{G}_{0g}(t \mid X_i^*)$. Then, $T_i^* = \min(Y_i^*, C_i^*) < \infty$ if $Y_i^* < \infty$ or $C_i^* < \infty$, that is with probability $1 - \left(1 - \widehat{p}_g^c(X_i^*)\right) \widehat{\pi}_g^c(X_i^*)$, and $T_i^* = C_{0i}^*$ if $Y_i^* = C_i^* = \infty$, which happens with probability $\left(1 - \widehat{p}_g^c(X_i^*)\right) \widehat{\pi}_g^c(X_i^*)$. The probability of T_i^* equals T_j, $j = 1, ..., n$ is as follows, depending on T_i^* corresponding to an individual known to be cured or not. First, let us consider the case of a generated individual unknown to be cured, that is, when $Y_i^* < \infty$ or $C_i^* < \infty$, then $T_i^* = \min(Y_i^*, C_i^*)$, $\delta_i^* = \mathbf{1}(Y_i^* < C_i^*)$ and $\xi_i^* \nu_i^* = \mathbf{1}(Y_i^* = \infty, C_i^* = \infty) = 0$. In this case, $T_i^* = \min(Y_i^*, C_i^*)$ is generated from the rescaled distribution function

$$\frac{1 - \left(1 - \widehat{F}_g^c(t \mid X_i^*)\right)\left(1 - \widehat{G}_g^c(t \mid X_i^*)\right)}{1 - \left(1 - \widehat{p}_g^c(X_i^*)\right)\widehat{\pi}_g^c(X_i^*)}.$$

Note that the sum of the weights given by $1 - \left(1 - \widehat{F}_g^c(t \mid X_i^*)\right)\left(1 - \widehat{G}_g^c(t \mid X_i^*)\right)$ is not

1 but $1-\left(1-\widehat{p}_g^c\left(X_i^*\right)\right)\widehat{\pi}_g^c\left(X_i^*\right)$.

From Proposition 2.2, the above distribution function of T_i^* reduces to

$$\frac{\widehat{H}_g^0\left(t \mid X_i^*\right)}{1-\left(1-\widehat{p}_g^c\left(X_i^*\right)\right)\widehat{\pi}_g^c\left(X_i^*\right)},$$

that puts mass only at the observed times T_j that are not classified as cured ($\xi_j\nu_j=0$). Then, $T_i^*=T_j$ with conditional probability

$$\begin{aligned}
&P\left(T_i^*=T_j \mid X_i^*\right)\\
&=P\left(T_i^*=T_j \mid Y_i^*<\infty \text{ or } C_i^*=\infty, X_i^*\right)\left(1-P\left(Y_i^*=\infty, C_i^*=\infty \mid X_i^*\right)\right)\\
&=\frac{B_{gj}\left(X_i^*\right)}{1-\left(1-\widehat{p}_g^c\left(X_i^*\right)\right)\widehat{\pi}_g^c\left(X_i^*\right)}\left[1-\left(1-\widehat{p}_g^c\left(X_i^*\right)\right)\widehat{\pi}_g^c\left(X_i^*\right)\right]\\
&=B_{gj}\left(X_i^*\right).
\end{aligned}$$

Now, we will prove that if $T_i^*=T_j$ then $\delta_i^*=\delta_j$ and $\xi_i^*\nu_i^*=\xi_j\nu_j$. Since $\xi_i^*\nu_i^*=\mathbf{1}\left(Y_i^*=\infty, C_i^*=\infty\right)=0$ and $\xi_j\nu_j=0$ then $\xi_i^*\nu_i^*=\xi_j\nu_j$. Regarding the bootstrap censoring indicator, recall that $1-\widehat{F}_g^c\left(t \mid X_i^*\right)$ puts mass only at those T_j with $\delta_j=1$ while $1-\widehat{G}_g^c\left(t \mid X_i^*\right)$ puts mass only at those T_j with $\delta_j=0$. Let $Y_i^*=T_k$ and $C_i^*=T_l$ be the bootstrapped lifetime and censoring time such that $T_i^*=\min(Y_i^*, C_i^*)$. Note that $\delta_k=1$ and $\delta_l=0$. Since we have assumed no ties between censored and uncensored times, if $\min\left(Y_i^*, C_i^*\right)=Y_i^*$ then $T_i^*=T_k$, $\delta_i^*=1$ and therefore $\delta_i^*=\delta_k$. While if $\min\left(Y_i^*, C_i^*\right)=C_i^*$ then $T_i^*=T_l$, $\delta_i^*=0$ and consequently $\delta_i^*=\delta_l$. In summary, let $(T_j, \delta_j, \xi_j\nu_j=0)$ be an individual in the subgroup with unknown cure status $\left\{\left(T_i, \delta_i, \xi_i\nu_i=0\right), i=1,\ldots,n-m\right\}$, then $(T_i^*, \delta_i^*, \xi_i^*\nu_i^*)$ equals $(T_j, \delta_j, \xi_j\nu_j=0)$ with probability $B_{gj}\left(X_i^*\right)$.

Let us consider now that the generated individual is known to be cured, that is $Y_i^*=C_i^*=\infty$. In this second case, T_i^* is a value generated from $\widehat{G}_{0g}\left(t \mid X_i^*\right)$ with:

$$\widehat{G}_{0g}\left(t \mid x\right)=\frac{\sum_{i=1}^n B_{gi}\left(x\right)\mathbf{1}\left(T_i\le t, \xi_i\nu_i=1\right)}{1-\sum_{i=1}^n B_{gi}\left(x\right)\mathbf{1}\left(\xi_i\nu_i=0\right)}=\frac{\sum_{i=1}^n B_{gi}\left(x\right)\mathbf{1}\left(T_i\le t, \xi_i\nu_i=1\right)}{1-\widehat{H}_g^0\left(T_{(n)} \mid x\right)}.$$

From Proposition 2.2,

$$1-\widehat{H}_g^0\left(T_{(n)} \mid x\right)=\left(1-\widehat{F}_g^c\left(T_{(n)} \mid x\right)\right)\left(1-\widehat{G}_g^c\left(T_{(n)} \mid x\right)\right)=\left(1-\widehat{p}_g^c\left(x\right)\right)\widehat{\pi}_g^c\left(x\right),$$

and, therefore,

$$\widehat{G}_{0g}\left(t \mid X_i^*\right) = \frac{\sum_{i=1}^n B_{gi}\left(X_i^*\right) \mathbf{1}\left(T_i \leq t, \xi_i\nu_i = 1\right)}{\left(1 - \widehat{p}_g^c\left(X_i^*\right)\right)\widehat{\pi}_g^c\left(X_i^*\right)}.$$

Note that $\widehat{G}_{0g}\left(t \mid X_i^*\right)$ puts mass only at those values T_j of a subject classified as cured ($\delta_j = 0$ and $\xi_j\nu_j = 1$). Let $\left(T_j, \delta_j, \xi_j\nu_j\right)$ be an individual in the subgroup of individuals known to be cured $\left\{\left(T_i, \delta_i = 0, \xi_i\nu_i = 1\right), i = n-m+1, \ldots, n\right\}$, then $T_i^* = T_j$ with probability

$$\begin{aligned} P\left(T_i^* = T_j \mid X_i^*\right) &= P\left(T_i^* = T_j \mid Y_i^* = \infty, C_i^* = \infty, X_i^*\right) P\left(Y_i^* = \infty, C_i^* = \infty \mid X_i^*\right) \\ &= \frac{B_{gj}\left(X_i^*\right)}{\left(1 - \widehat{p}_g^c\left(X_i^*\right)\right)\widehat{\pi}_g^c\left(X_i^*\right)}\left(1 - \widehat{p}_g^c\left(X_i^*\right)\right)\widehat{\pi}_g^c\left(X_i^*\right) = B_{gj}\left(X_i^*\right). \end{aligned}$$

Again, we will show that if $T_i^* = T_j$ when T_j corresponds to an individual known to be cured ($\delta_j = 0$, $\xi_j\nu_j = 1$) then $\delta_i^* = \delta_j$ and $\xi_i^*\nu_i^* = \xi_j\nu_j$.

In this second case, $Y_i^* = C_i^* = \infty$, which yields $\delta_i^* = \mathbf{1}\left(Y_i^* < C_i^*\right) = 0$ and $\xi_i^*\nu_i^* = \mathbf{1}\left(Y_i^* = \infty, C_i^* = \infty\right) = 1$. In summary, when $\left(T_j, \delta_j, \xi_j\nu_j\right)$ is an individual known to be cured, then $\left(T_i^*, \delta_i^*, \xi_i^*\nu_i^*\right)$ equals $\left(T_j, \delta_j, \xi_j\nu_j\right)$ with probability $B_{gj}\left(X_i^*\right)$. This completes the proof. □

Appendix B

Proofs of the results in Chapter 3

Proof of Proposition 3.1

Proposition 3.1 The estimator $1 - \widehat{p}_h^c(x)$ has the following properties.

1. When there are no censored observations known to be cured, i.e., $\xi_i \nu_i = 0$ for $i = 1, \dots, n$, $1 - \widehat{p}_h^c(x)$ reduces to the XP estimator in (1.5).

   Proof. It is straightforward since $\xi_i \nu_i = 0$, $i = 1, \dots, n$. □

2. In the specific case that some individuals are classified as cured when their survival time exceeds a known fixed cure threshold, $1 - \widehat{p}_h^c(x)$ also reduces to the XP estimator.

   Proof. Assume there exists a common specific known cure threshold $d_i = d$ for $i = 1, \dots, n$. This implies that in the ordered sample, $\left\{ \left(X_{[i]}, T_{(i)}, \delta_{[i]}, \xi_{[i]} \nu_{[i]} \right), \right.$ $\left. i = 1, \dots, n \right\}$, the n_1 first observations correspond to individuals with $T_{(i)} < d$ either not cured or with unknown cure status ($\xi_{[i]} \nu_{[i]} = 0$), and the remaining m observations are from cured individuals with $T_{(i)} \geq d$ and $\xi_{[i]} \nu_{[i]} = 1$. Therefore,

$$\begin{aligned} 1 - \widehat{p}_h^c(x) &= \prod_{i=1}^{n} \left(1 - \frac{\delta_{[i]} B_{h[i]}(x)}{\sum_{j=i}^{n_1} B_{h[j]}(x) + \sum_{j=n_1+1}^{n} B_{h[j]}(x)} \right) \\ &= \prod_{i=1}^{n} \left(1 - \frac{\delta_{[i]} B_{h[i]}(x)}{\sum_{j=i}^{n} B_{h[j]}(x)} \right) = 1 - \widehat{p}_h(x). \end{aligned}$$

   □

3. When there is no censoring, all the cure status indicators ν_i are observed ($\xi_i = 1, i = 1, \dots, n$). In this case, $1 - \widehat{p}_h^c(x)$ reduces to the NW estimator of the cure probability

Proof. Without censoring, $T_i = Y_i, \delta_i = 1$ and the cure status is always observed $\xi_i = 1$. In this situation, the $n = n_1 + m$ observations can be ordered and split into the n_1 uncured individuals with finite lifetimes Y_i, and the m cured individuals with lifetime $Y_i = \infty$. Thus,

$$\begin{aligned}
1 - \widehat{p}_h^c(x) &= \prod_{i=1}^{n} \left(1 - \frac{B_{h[i]}(x)}{\sum_{j=i}^{n} B_{h[j]}(x) + \sum_{j=1}^{i-1} B_{h[j]}(x) \mathbf{1}\left(\nu_{[j]} = 1\right)} \right) \\
&= \prod_{i=1}^{n} \left(\frac{\sum_{j=i+1}^{n} B_{h[j]}(x) + \sum_{j=1}^{i} B_{h[j]}(x) \mathbf{1}\left(\nu_{[j]} = 1\right)}{\sum_{j=i}^{n} B_{h[j]}(x) + \sum_{j=1}^{i-1} B_{h[j]}(x) \mathbf{1}\left(\nu_{[j]} = 1\right)} \right) \\
&= \frac{\sum_{j=2}^{n} B_{h[j]}(x) + B_{h[j]}(x) \mathbf{1}\left(\nu_{[j]} = 1\right)}{\sum_{j=1}^{n} B_{h[j]}(x)} \\
&\quad \times \frac{\sum_{j=3}^{n} B_{h[j]}(x) + \sum_{j=1}^{2} B_{h[j]}(x) \mathbf{1}\left(\nu_{[j]} = 1\right)}{\sum_{j=2}^{n} B_{h[j]}(x) + B_{h[j]}(x) \mathbf{1}\left(\nu_{[j]} = 1\right)} \\
&\quad \times \ldots \times \frac{\sum_{j=1}^{n} B_{h[j]}(x) \mathbf{1}\left(\nu_{[j]} = 1\right)}{B_{h[n]}(x) + \sum_{j=1}^{n-1} B_{h[j]}(x) \mathbf{1}\left(\nu_{[j]} = 1\right)} \\
&= \frac{\sum_{j=1}^{n} B_{h[j]}(x) \mathbf{1}\left(\nu_{[j]} = 1\right)}{\sum_{j=1}^{n} B_{h[j]}(x)} = \sum_{j=1}^{n} B_{h[j]}(x) \mathbf{1}\left(\nu_{[j]} = 1\right).
\end{aligned}$$

□

4. In an unconditional setting, the proposed estimator is

$$1 - \widehat{p}_n^c = \prod_{i=1}^{n} \left(1 - \frac{\delta_{[i]}}{n - i + 1 + \sum_{j=1}^{i-1} \mathbf{1}\left(\xi_{[j]} \nu_{[j]} = 1\right)} \right).$$

Proof. In an unconditional setting the weights are $1/n$ for $i = 1, \ldots, n$. Thus, the $1 - \widehat{p}_h^c(x)$ becomes

$$1 - \widehat{p}_n^c = \prod_{i=1}^{n} \left(1 - \frac{\delta_{[i]} n^{-1}}{n^{-1}(n - i + 1) + n^{-1} \sum_{j=1}^{i-1} \mathbf{1}\left(\xi_{[j]} \nu_{[j]} = 1\right)} \right).$$

In the particular case where an individual is known to be cured only if the observed time is greater than a known fixed time, say d, with $n = n_1 + m$ observations, where m are identified as cured, the ordered observed lifetimes $T_{(1)} \leq \ldots \leq T_{(n_1)}$ are strictly lower than d, and the m cured individuals have $T_{(i)} \geq d$. Besides, the weights are $1/n$ for $i = 1, \ldots, n$. Then the proposed estimator reduces to the one in Laska and Meisner (1992)

$$1-\widehat{p}_n^c=\prod_{i=1}^{n}\left(1-\frac{\delta_{[i]}}{n_1-i+1+m}\right).$$

□

Proof of Proposition 3.2

Proposition 3.2 The estimator $1-\widehat{p}_h^c\left(x\right)$ in (3.1) is the nonparametric local maximum likelihood estimator of $1-p(x)$.

Proof. Mimicking the proof of Proposition 2.3 in Appendix A, it is not difficult to show that the estimator $1-\widehat{p}_h^c\left(x\right)$ is the local maximum likelihood estimator of $1-p(x)$. □

Proof of Theorem 3.1

Theorem 3.1 Suppose that Assumptions $1-9$ and condition (3.5) hold, then for $x\in I$ one obtains

$$(1-\widehat{p}_h^c(x))-(1-p(x))=(1-p(x))\sum_{i=1}^{n}\widetilde{B}_{hi}\left(x\right)\zeta\left(T_i,\delta_i,\xi_i,\nu_i,\tau_0,x\right)+R_n\left(x\right),$$

where $\zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right)$ is given in (2.21), $\widetilde{B}_{hi}\left(x\right)$ is defined in (2.22), and $R_n\left(x\right)$ satisfies

$$\sup_{x\in I}\mid R_n\left(x\right)\mid=O\left(\left(\log n\right)^{3/4}\left(nh\right)^{-3/4}\right)\text{ a.s.}$$

Proof. Consider the following decomposition

$$\begin{aligned}(1-\widehat{p}_h^c(x))-(1-p(x))=&\left(1-\widehat{F}_h^c\left(T_{(n)}^1\mid x\right)\right)-\left(1-F\left(\tau_0\mid x\right)\right)\\&\pm\exp\left(-\widehat{\Lambda}_h^c\left(T_{(n)}^1\mid x\right)\right)\\=&\exp\left(-\widehat{\Lambda}_h^c\left(T_{(n)}^1\mid x\right)\right)-\left(1-F\left(\tau_0\mid x\right)\right)+R_1(x)\qquad\text{(B.1)}\end{aligned}$$

where $R_1\left(x\right)=\left(1-\widehat{F}_h^c\left(T_{(n)}^1\mid x\right)\right)-\exp\left(-\widehat{\Lambda}_h^c\left(T_{(n)}^1\mid x\right)\right)$, and $\widehat{\Lambda}_h^c\left(t\mid x\right)$ is given in (2.15).

Note that by a Taylor's expansion of the exponential function around $-\Lambda(\tau_0\mid x)$, (B.1) becomes

$$(1-\widehat{p}_h^c(x))-(1-p(x))=-\left(1-p\left(x\right)\right)\left(\widehat{\Lambda}_h^c\left(T_{(n)}^1\mid x\right)-\Lambda\left(\tau_0\mid x\right)\right)+R_1\left(x\right)$$

$$+ R_2(x), \tag{B.2}$$

where $R_2(x) = -\frac{1}{2}\exp(-\Lambda(t^* \mid x))\left(\widehat{\Lambda}_h^c\left(T_{(n)}^1 \mid x\right) - \Lambda(\tau_0 \mid x)\right)^2$ and $\Lambda(t^* \mid x)$ is a value between $\widehat{\Lambda}_h^c\left(T_{(n)}^1 \mid x\right)$ and $\Lambda(\tau_0 \mid x)$. Further, (B.2) can be decomposed as

$$\begin{aligned}
&(1-\widehat{p}_h^c(x)) - (1-p(x)) \\
&= -\left(\widehat{\Lambda}_h^c\left(T_{(n)}^1 \mid x\right) - \Lambda(\tau_0 \mid x) \pm \widehat{\Lambda}_h^c(\tau_0 \mid x)\right)(1-p(x)) + R_1(x) + R_2(x) \\
&= -\left(\widehat{\Lambda}_h^c(\tau_0 \mid x) - \Lambda(\tau_0 \mid x)\right)(1-p(x)) + R_1(x) + R_2(x) + R_3(x)
\end{aligned}$$

where

$$R_3(x) = \ -(1-p(x))\left(\widehat{\Lambda}_h^c\left(T_{(n)}^1 \mid x\right) - \widehat{\Lambda}_h^c(\tau_0 \mid x)\right).$$

Arguing similarly as in the proof of Theorem 2 in Iglesias-Pérez and González-Manteiga (1999), given $t = T_{(n)}^1$ under Assumptions 1,–9, then

$$\sup_{x\in I} \mid R_1(x) \mid = O\left(n^{-1}h^{-1}\right) \text{ a.s.}$$

For the term $R_2(x)$, Lemma 5 in López-Cheda et al. (2017a) under Assumption 7 is used:

$$n^{\alpha}\left(\tau_0 - T_{(n)}^1\right) \to 0 \text{ a.s. for any } \alpha \in (0,1), \tag{B.3}$$

and the strong consistency results for the estimator $\widehat{\Lambda}_h^c(t \mid x)$ in Corollary 2.1 for $t = \tau_0$. Then

$$\sup_{x\in I} \mid R_2(x) \mid = O\left(n^{-1}h^{-1}\log n\right) \text{ a.s.}$$

The third term is bounded as follows,

$$\sup_{x\in I} \mid R_3(x) \mid \leq \mid T_{(n)}^1 - \tau_0 \mid \widehat{\Lambda}_h^c(t^* \mid x)(1-p(x)),$$

where $\widehat{\Lambda}_h^c(t^* \mid x)$ is a value between $\widehat{\Lambda}_h^c\left(T_{(n)}^1 \mid x\right)$ and $\widehat{\Lambda}_h^c(\tau_0 \mid x)$. From (B.3) for a sequence of bandwidths satisfying $h \to 0$, one obtains

$$\tau_0 - T_{(n)}^1 = O\left((\log n)^{3/4}(nh)^{-3/4}\right) \text{ a.s.}$$

and as a consequence

$$\sup_{x\in I} \mid R_3(x) \mid = O\left((\log n)^{3/4}(nh)^{-3/4}\right) \quad \text{a.s.}$$

The proof concludes by applying the results in Theorem 2.1. □

Proof of Proposition 3.3

Proposition 3.3 Suppose that Assumptions $1-9$ and condition (3.5) hold, then, the asymptotic bias and variance of the dominant term of $1-\widehat{p}_h^c\left(x\right)$ are, respectively,

$$\mu_{h,c}(x) = h^2 B_c(x) + O\left(h^4\right) \quad \text{and} \quad \sigma^2_{h,c}(x) = \frac{1}{nh} s_c^2(x) + O\left(\frac{h}{n}\right),$$

where $B_c(x)$ in the dominant term of the bias is

$$B_c(x) = \left(c_{1,c}(x) + c_{2,c}(x)\right) d_K$$

with $d_K = \int v^2 K(v) dv$,

$$\begin{aligned} c_{1,c}(x) &= \frac{2(1-p(x))' m'(x) + (1-p(x))'' m(x)}{2m(x)}, \\ c_{2,c}(x) &= (1-p(x)) \int_0^{\tau_0} \frac{G'(v^- \mid x)}{1-G(v^- \mid x)} \frac{d}{ds}\left(\frac{S'\left(s \mid x\right)}{S\left(s \mid x\right)}\right)|_{s=v^-} dv. \end{aligned}$$

Here $p'(x), p''(x), S'(t \mid x)$ and $G'(t \mid x)$ refer to the derivatives with respect to x. The function $s_c^2(x)$ in the dominant term of the variance is

$$s_c^2(x) = \frac{\left(1-p\left(x\right)\right)^2}{m\left(x\right)} \int_0^{\tau_0} \frac{dH^1\left(v^- \mid x\right)}{\left(1-H\left(v^- \mid x\right) + H^{11}(v^- \mid x)\right)^2} c_K,$$

with $c_K = \int K^2(v) dv$.

Proof. From Theorem 3.1, the asymptotic bias of $1-\widehat{p}_h^c\left(x\right)$ equals to expected value of

$$\frac{(nh)^{-1}(1-p\left(x\right))}{m\left(x\right)} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \zeta\left(T_i, \delta_i, \xi_i, \nu_i, \tau_0, x\right) = I + II, \tag{B.4}$$

where

$$\begin{aligned} I = \frac{(nh)^{-1}(1-p\left(x\right))}{m\left(x\right)} & \left[\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \zeta\left(T_i, \delta_i, \xi_i, \nu_i, \tau_0, x\right)\right. \\ & \left. - E\left(\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \zeta\left(T_i, \delta_i, \xi_i, \nu_i, \tau_0, x\right)\right)\right], \end{aligned} \tag{B.5}$$

$$II = \frac{(nh)^{-1}(1-p\left(x\right))}{m\left(x\right)} E\left(\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \zeta\left(T_i, \delta_i, \xi_i, \nu_i, \tau_0, x\right)\right), \tag{B.6}$$

with $\zeta(T,\delta,\xi,\nu,t,x)$ in (2.21).

Since $E(I)=0$, the asymptotic bias of the estimator $1-\widehat{p}_h^c(x)$ is II. Using Lemmas D.1 and D.2 when $t=\tau_0$, then one obtains

$$II=\frac{h^2(1-p(x))(\Phi_c''(x,\tau_0,x)\,m(x)+2\Phi_c'(x,\tau_0,x)\,m'(x))d_K}{2m(x)}+O\left(h^4\right), \tag{B.7}$$

with $\Phi_c(y,t,x)=E(\zeta(T,\delta,\xi,\nu,t,x)\mid X=y)$, and $\Phi_c'(y,t,x)$ and $\Phi_c''(y,t,x)$ are the first and the second derivatives of $\Phi_c(y,t,x)$ with respect to y.

By applying Lemma D.3 for $t=\tau_0$, it suffices to show that

$$\Phi_c'(x,\tau_0,x)=-\frac{(1-p(x))'}{1-p(x)}. \tag{B.8}$$

Besides, because of Lemma D.4 one can show that

$$\Phi_c''(x,\tau_0,x)=2\int_0^{\tau_0}\frac{G'(v^-\mid x)}{1-G(v^-\mid x)}\frac{d}{ds}\left(\frac{S'(s\mid x)}{S(s\mid x)}\right)|_{s=v^-}dv-\frac{(1-p(x))''}{1-p(x)}, \tag{B.9}$$

where $G(t\mid x)=(1-\pi(x))G_0(t\mid x)$. The expression of the asymptotic bias of $1-\widehat{p}_h^c(x)$ derives from plugging (B.8) and (B.9) in (B.7).

Recalling (B.4), the asymptotic variance of $1-\widehat{p}_h^c(x)$ is

$$\operatorname{Var}(I)=\frac{(1-p(x))^2}{m^2(x)}(V_1-V_2), \tag{B.10}$$

where

$$\begin{aligned} V_1&=\frac{1}{nh^2}E\left(K^2\left(\frac{x-X}{h}\right)\zeta^2(T,\delta,\xi,\nu,\tau_0,x)\right),\\ V_2&=\frac{1}{nh^2}\left[E\left(K\left(\frac{x-X}{h}\right)\zeta(T,\delta,\xi,\nu,\tau_0,x)\right)\right]^2. \end{aligned}$$

From Lemmas D.1 and D.1 for $t=\tau_0$, V_2 reduces to

$$\begin{aligned} V_2&=\frac{1}{4}\frac{h^2}{n}d_K^2\left(\frac{\Phi_c''(x,\tau_0,x)\,m(x)+2\Phi_c'(x,\tau_0,x)\,m'(x)}{m(x)}\right)^2+O\left(n^{-1}h^4\right)\\ &=O\left(n^{-1}h^2\right). \end{aligned} \tag{B.11}$$

As for V_1, recall $\Phi_1^c(y,\tau_0,x)=E\left(\zeta^2(T,\delta,\xi,\nu,\tau_0,x)\mid X=y\right)$. Then, after a change of variable and a Taylor's expansion one obtains

$$V_1=\frac{1}{nh}\Phi_1^c(x,\tau_0,x)\,m(x)\,c_K+O\left(n^{-1}h\right). \tag{B.12}$$

From Lemma D.5, the function $\Phi_1^c(x,t,x)$ can be written as:

$$\Phi_1^c\left(x,t,x\right) = \int_0^t \frac{dH^1\left(v^- \mid x\right)}{\left(1 - H\left(v^- \mid x\right) + H^{11}(v^- \mid x)\right)^2}. \tag{B.13}$$

The proof concludes by substituting (B.11) and (B.12) into (B.10) and using (B.13). □

Proof of Theorem 3.2

Theorem 3.2 Suppose that Assumptions 1 − 9 and condition (3.5) are satisfied, then for $x \in I$ it follows that:

(i) If $nh^5 \to 0$ and $(\log n)^3/(nh) \to 0$, then

$$(nh)^{1/2}\left(\widehat{p}_h^c(x) - p(x)\right) \xrightarrow{d} N(0, s_c^2(x)).$$

(ii) If $nh^5 \to C$, where $C > 0$ is a constant then

$$(nh)^{1/2}\left(\widehat{p}_h^c(x) - p(x)\right) \xrightarrow{d} N(C^{5/2}B_c(x), s_c^2(x)),$$

where $B_c(x)$ is defined in (3.7) and $s_c^2(x)$ in (3.10).

Proof. From Theorem 3.1, let us consider

$$\begin{aligned}&(nh)^{1/2}[(1 - \widehat{p}_h^c\left(x\right)) - (1 - p\left(x\right))]\\ &= (nh)^{1/2}\left[(1 - p\left(x\right))\sum_{i=1}^n \widetilde{B}_{hi}\left(x\right)\zeta\left(T_i, \delta_i, \xi_i, \nu_i, \tau_0, x\right) + R_n\left(x\right)\right],\end{aligned}$$

with $\zeta(T, \delta, \xi, \nu, t, x)$ and $R_n(x)$ given in (2.21) and (3.6), respectively. The condition $(\log n)^3/(nh) \to 0$ implies that $(nh)^{1/2}(\log n/(nh))^{3/4} \to 0$, so the remainder term $(nh)^{1/2}R_n(x)$ is negligible. Consequently, the asymptotic distribution of $(nh)^{1/2}[(1 - \widehat{p}_h^c\left(x\right)) - (1 - p\left(x\right))]$ is that of

$$\begin{aligned}&(nh)^{1/2}\frac{1 - p\left(x\right)}{m\left(x\right)}\sum_{i=1}^n \frac{1}{nh}K\left(\frac{x - X_i}{h}\right)\zeta\left(T_i, \delta_i, \xi_i, \nu_i, \tau_0, x\right)\\ &= (nh)^{1/2}(I + II),\end{aligned} \tag{B.14}$$

where I and II are given in (B.5) and (B.6). Under the assumption $nh^5 \to 0$, one has $(nh)^{1/2}II = o(1)$. Therefore, the asymptotic distribution of (B.14) is that of $(nh)^{1/2}I$.

Let $(nh)^{1/2}I = \sum_{i=1}^{n} \eta_{i,h}(x)$, where

$$\eta_{i,h}(x) = \frac{(nh)^{-1/2}(1-p\left(x\right))}{m\left(x\right)}\left[K\left(\frac{x-X_i}{h}\right)\zeta\left(T_i,\delta_i,\xi_i,\nu_i,\tau_0,x\right)\right.$$
$$\left. - E\left(K\left(\frac{x-X_i}{h}\right)\zeta\left(T_i,\delta_i,\xi_i,\nu_i,\tau_0,x\right)\right)\right],$$

Lindeberg's theorem for triangular arrays (Billingsley, 1968) can be applied to obtain

$$\frac{\sum\limits_{i=i}^{n}\eta_{i,h}\left(x\right)}{s_c^2(x)} \to N\left(0,1\right) \quad \text{in distribution,}$$

with $s_c^2(x)$ in (3.10). Therefore, $(nh)^{1/2}[(1-\widehat{p}_h^c\left(x\right)) - (1-p\left(x\right))] \to N(0, s_c^2(x))$ in distribution. This proves (i). The proof of (ii) is similar, noting that if $nh^5 = C^5$, then the bias term is non-negligible. □

Proof of Proposition 3.5

Proposition 3.5 Suppose that the siMAR condition and Assumptions 1 (i), 2 (i), 3 (i), 8, 10 hold. Also, the bandwidths h, g_1, g_2 satisfy $h \to 0$, $g_1 \to 0$, $g_2 \to 0$, $nh \to \infty$, $ng_1 \to \infty$ and $ng_2 \to \infty$ as $n \to \infty$. The asymptotic bias of $1 - \widehat{p}_h^{\text{MI-NW}}(x)$ is

$$\mu_{g_1,g_2,h}^{\text{MI-NW}}(x) = h^2 c_{1,c}(x) + \left(g_1^2 + g_2^2\right)c_{2,\text{MI-NW}}(x) + o\left((h^2+g_1^2+g_2^2)^2\right),$$

where $c_{1,c}(x)$ is defined in (3.8), and

$$c_{2,\text{MI-NW}}(x) = \frac{(1-\pi\left(x\right))\left[\pi\left(x\right)\left(1-p\left(x\right)\right)m\left(x\right)\right]''}{2m\left(x\right)\pi\left(x\right)}d_K.$$

If the bandwidths are $g_1/h \to C_1$ and $g_2/h \to C_2$, then the asymptotic variance is

$$\begin{aligned}\sigma^2_{h,\text{MI-NW}}\left(x\right) = \frac{1}{nh}\frac{1-p(x)}{m(x)}\Bigg(&\frac{c_K(1-\pi\left(x\right))p\left(x\right)}{M\pi\left(x\right)} \\ &+\left\{\pi(x)c_K + (1-\pi(x))\left[c_{K,C_1,C_2} + \frac{1-\pi(x)}{\pi(x)}d_{K,C_1,C_2}\right.\right. \\ &\left.\left.+\ (1-p(x))\left(c_K + 2c_{K,C_2} + \frac{1-\pi(x)}{\pi(x)}(c_{K,C_1,C_2} + 2d_{K,C_1,C_2})\right)\right]\right\}\Bigg) \\ &+\frac{2}{ng_1}\left(1-p\left(x\right)\right)^2\frac{1-\pi\left(x\right)}{\pi\left(x\right)}K\left(0\right) + o\left((Mnh)^{-1}\right) + o\left((nh)^{-1}\right) \\ &+ o\left((ng_1)^{-1}\right),\end{aligned}$$

where $c_{K,C} = \iint K(u)K(v)K(u+Cv)dudv$,

$$c_{K,C_1,C_2} = \iiint K(u)K(v)K(w)K(u+C_1v+C_2w)dudvdw$$

and

$$d_{K,C_1,C_2} = \iiint K(u)K(v)K(w)K(u+C_1v+C_2(u+w))dudvdw.$$

Proof. Observe that the expectation of $1-\widehat{p}_h^{\text{MI-NW}}(x)$ is

$$\begin{aligned} E\left(1-\widehat{p}_h^{\text{MI-NW}}(x)\right) &= \frac{1}{m}\sum_{m=1}^{M}\sum_{i=1}^{n} E\left(B_{hi}(x)\,\widetilde{\nu}_i^m\right) \\ &= \sum_{i=1}^{n} E\left(B_{hi}(x)\,\xi_i\nu_i\right) + \frac{1}{m}\sum_{m=1}^{M}\sum_{i=1}^{n} E\left(B_{hi}(x)\,(1-\xi_i)\,\nu_i^{+,m}\right) \\ &= E_1(x) + E_2(x). \end{aligned} \tag{B.15}$$

The NW weights verify

$$\begin{aligned} B_{hi}(x) &= \frac{K_h(x-X_i)}{\sum_{j=1}^{n} K_h(x-X_j)} = \frac{K_h(x-X_i)}{n\widehat{m}_h(x)} \\ &= \frac{K_h(x-X_i)}{n\widehat{m}_h(x)}\left[\frac{\widehat{m}_h(x)}{m(x)} + \frac{m(x)-\widehat{m}_h(x)}{m(x)}\left(\frac{\widehat{m}_h(x)}{m(x)} + \frac{m(x)-\widehat{m}_h(x)}{m(x)}\right)\right] \\ &= \frac{K_h(x-X_i)}{nm(x)} + \frac{K_h(x-X_i)}{nm^2(x)}\left(m(x)-\widehat{m}_h(x)\right) \\ &\quad + \frac{K_h(x-X_i)}{nm(x)}\frac{\left(m(x)-\widehat{m}_h(x)\right)^2}{\widehat{m}_h(x)} \\ &= \frac{K_h(x-X_i)}{nm(x)}\left(2-\frac{1}{nm(x)}\sum_{j=1}^{n} K_h(x-X_j)\right) \\ &\quad + O\left(n^{-1}\left(h^2+(nh)^{-1/2}(\log n)^{1/2}\right)^2\right) \quad \text{a.s.} \end{aligned}$$

as $n\to\infty$ under Assumption 1(i), where $\widehat{m}_h(x)$ is the kernel estimator (Rosenblatt, 1956; Parzen, 1962) of the density function $m(x)$. Then, for the summands in $E_1(x)$ one has

$$\begin{aligned} &E\left(B_{hi}(x)\,\xi_i\nu_i\right) \\ &= \frac{1}{nm(x)}\left[2E\left(K_h(x-X_i)\,\xi_i\nu_i\right) - \frac{1}{nm(x)}E\left(K_h(x-X_i)\sum_{j=1}^{n} K_h(x-X_j)\,\xi_i\nu_i\right)\right] \\ &\quad + O\left(n^{-1}\left(h^2+(nh)^{-1/2}(\log n)^{1/2}\right)^2\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{nm(x)}\Bigg[2E\left(K_h(x-X_i)\,\xi_i\nu_i\right) - \frac{1}{nm(x)}E\left(K_h^2(x-X_i)\,\xi_i\nu_i\right) \\
&\quad + E\left(K_h(x-X_i)\sum_{\substack{j=1\\ j\neq i}}^{n} K_h(x-X_j)\,\xi_i\nu_i\right)\Bigg] + O\left(n^{-1}\left(h^2+(nh)^{-1/2}(\log n)^{1/2}\right)^2\right) \\
&= \frac{1}{nm(x)}\Bigg\{2E\left[K_h(x-X_i)\,E\left(\xi_i\nu_i \mid X_i\right)\right] - \frac{1}{nm(x)}E\left[K_h^2(x-X_i)\,E\left(\xi_i\nu_i \mid X_i\right)\right] \\
&\quad - E\left[K_h(x-X_i)\sum_{\substack{j=1\\ j\neq i}}^{n} K_h(x-X_j)\,E\left(\xi_i\nu_i \mid X_i\right)\right]\Bigg\} \\
&\quad + O\left(n^{-1}\left(h^2+(nh)^{-1/2}(\log n)^{1/2}\right)^2\right)
\end{aligned}
$$

Under the siMAR assumption,

$$
E\left(\xi_i\nu_i \mid X_i\right) = \pi\left(X_i\right)\left(1-p\left(X_i\right)\right). \tag{B.16}
$$

Thus, using (B.16), applying Lemmas D.6 – D.8 the first expectation in (B.15) can be derived as

$$
\begin{aligned}
E_1(x) &= \frac{1}{m(x)}\Bigg\{2E\left[K_h(x-X_1)\,\pi(X_1)\left(1-p(X_1)\right)\right] \\
&\quad - \frac{1}{nm(x)}E\left[K_h^2(x-X_1)\,\pi(X_1)\left(1-p(X_1)\right)\right] \\
&\quad - E\left[K_h(x-X_1)\,K_h(x-X_2)\,\pi(X_1)\left(1-p(X_1)\right)\right]\Bigg\} + O\left(n^{-1}\left(h^4+n^{-1}h^{-1}\right)\right) \\
&= \pi(x)\left(1-p(x)\right)\left(2-\frac{n-1}{n}\right) + h^2\frac{\left(\pi(x)\left(1-p(x)\right)m(x)\right)''(x)}{m(x)}d_K \\
&\quad - \frac{h^2}{2}\frac{n-1}{n}\frac{\left[\pi(x)\left(1-p(x)\right)m''(x) + \left(\pi(x)\left(1-p(x)\right)m(x)\right)''(x)\right]}{m(x)}d_K \\
&\quad - \frac{1}{nh}\frac{\pi(x)\left(1-p(x)\right)}{m(x)}c_K + O\left(n^{-1}\left(h^2+(nh)^{-1/2}(\log n)^{1/2}\right)^2\right) \\
&= \frac{n+1}{n}\pi(x)\left(1-p(x)\right) + \frac{1}{2}h^2\frac{1}{m(x)}\Bigg\{\frac{n+1}{n}\left[\pi(x)\left(1-p(x)\right)m(x)\right]'' \\
&\quad - \frac{n-1}{n}\pi(x)\left(1-p(x)\right)m''(x)\Bigg\}d_K + O\left(\left(h^2+(nh)^{-1/2}(\log n)^{1/2}\right)^2\right).
\end{aligned} \tag{B.17}
$$

The terms of $E_2(x)$ are

$$\begin{aligned} E\left(B_{hi}\left(x\right)\left(1-\xi_i\right)\nu_i^{+,m}\right) &= E\left(B_{hi}\left(x\right)\left(1-\xi_i\right)E\left(\nu_i^{+,m}|\text{observed data}\right)\right) \\ &= E\left(B_{hi}\left(x\right)\left(1-\xi_i\right)\sum_{k=1}^{n}B_{g_1k}^{\xi}(X_i)\sum_{j=1}^{n}B_{g_2j}^{\xi}(X_k)\nu_j\right), \end{aligned}$$

where

$$B_{gj}^{\xi}\left(x\right)=\frac{\xi_jK_g\left(x-X_j\right)}{\sum_{k=1}^{n}\xi_kK_g\left(x-X_k\right)} \tag{B.18}$$

are the kernel weights used in Steps 1 and 2 of Section 3.3.2.

By applying Lemma D.6, the denominator in (B.18) can be expanded as

$$\begin{aligned} &E\left(\frac{1}{n}\sum_{k=1}^{n}\xi_kK_g\left(x-X_k\right)\right) \\ &=\frac{n-1}{n}\left(\pi\left(x\right)m\left(x\right)+\frac{g^2}{2}\pi\left(x\right)m''\left(x\right)d_K+O\left(g^4\right)\right)+O\left(\frac{1}{ng}\right). \end{aligned}$$

It can be verified that

$$\begin{aligned} B_{gj}^{\xi}\left(x\right) &= \frac{\xi_jK_g\left(x-X_j\right)}{\sum_{k=1}^{n}\xi_kK_g\left(x-X_k\right)} \\ &= \frac{\xi_jK\left(\frac{x-X_j}{g}\right)}{ng\pi\left(x\right)m\left(x\right)}\left(1+\frac{ng\pi\left(x\right)m\left(x\right)-\sum_{k=1}^{n}\xi_kK\left(\frac{x-X_k}{g}\right)}{\sum_{k=1}^{n}\xi_kK\left(\frac{x-X_k}{g}\right)}\right), \\ &\simeq \frac{1}{ng}\frac{\xi_jK\left(\frac{x-X_j}{g}\right)}{\pi\left(x\right)m\left(x\right)}\left(1+o\left(1\right)\right). \end{aligned}$$

Therefore, the following expressions can be obtained after straightforward calculations

$$\begin{aligned} E\left(B_{hi}\left(x\right)\left(1-\xi_i\right)\nu_i^{+,m}\right) \simeq \frac{1}{n^2g_1g_2}\sum_{k=1}^{n}\sum_{j=1}^{n}E\Big[&\frac{\left(\pi\left(X_i\right)\pi\left(X_k\right)\right)^{-1}}{m\left(X_i\right)m\left(X_k\right)}K\left(\frac{X_i-X_k}{g_1}\right) \\ &\times K\left(\frac{X_j-X_k}{g_2}\right)B_{hi}\left(x\right)\left(1-\xi_i\right)\xi_j\xi_k\nu_j\Big]. \end{aligned}$$

Note that $\left(1-\xi_i\right)\xi_k\xi_j=0$ if $i=j$ or $i=k$. So, there are two cases to be considered: (a) $i\neq j$, $i\neq k$, $k=j$, and (b) $i\neq j$, $i\neq k$, $k\neq j$. For case (a) it can be shown

$$\begin{aligned} &E\left(B_{hi}\left(x\right)\left(1-\xi_i\right)\nu_i^{+,m}\right) \\ &=\frac{1}{n^2g_1g_2}\sum_{\substack{j=1 \\ j\neq i,k=j}}^{n}E\Big(\frac{\left(\pi\left(X_i\right)\pi\left(X_k\right)\right)^{-1}}{m\left(X_i\right)m\left(X_k\right)}K\left(\frac{X_i-X_k}{g_1}\right)K\left(\frac{X_j-X_k}{g_2}\right) \end{aligned}$$

$$
\begin{aligned}
&\times B_{hi}(x)(1-\xi_i)\xi_k\xi_j\nu_j\Big)\\
&= \frac{1}{n^2 g_1 g_2} K(0) \sum_{\substack{j=1\\ j\neq i}}^{n} E\left(\frac{(\pi(X_i)\pi(X_j))^{-1}}{m(X_i)m(X_j)} K\left(\frac{X_i-X_j}{g_1}\right) B_{hi}(x)(1-\xi_i)\xi_j\nu_j\right)\\
&\simeq \frac{1}{n^2 g_1 g_2}\frac{1}{nh}\frac{1}{m(x)} K(0) \sum_{\substack{j=1\\ j\neq i}}^{n} E\Bigg(\frac{(\pi(X_i)\pi(X_j))^{-1}}{m(X_i)m(X_j)} K\left(\frac{X_i-X_j}{g_1}\right)\\
&\times K\left(\frac{x-X_i}{h}\right)(1-\xi_i)\xi_j\nu_j\Bigg)\\
&= \frac{1}{n^2 g_1 g_2}\frac{1}{nh}\frac{1}{m(x)} K(0) \sum_{\substack{j=1\\ j\neq i}}^{n} E\Bigg[\frac{(\pi(X_i)\pi(X_j))^{-1}}{m(X_i)m(X_j)} K\left(\frac{X_i-X_j}{g_1}\right) K\left(\frac{x-X_i}{h}\right)\\
&\times E\left((1-\xi_i)\xi_j\nu_j \mid X_i, X_j, X_k\right)\Bigg]\\
&= \frac{1}{n^2 g_1 g_2}\frac{1}{nh}\frac{1}{m(x)} K(0) \sum_{\substack{j=1\\ j\neq i}}^{n} E\Bigg(K\left(\frac{X_i-X_j}{g_1}\right) K\left(\frac{x-X_i}{h}\right)\frac{1-\pi(X_i)}{\pi(X_i)m(X_i)}\\
&\times \frac{1-p(X_j)}{m(X_j)}\Bigg)\\
&= \frac{1}{n^2 g_1 g_2}\frac{n-1}{nh}\frac{1}{m(x)} K(0)\, E\Bigg(K\left(\frac{X_1-X_2}{g_1}\right) K\left(\frac{x-X_1}{h}\right)\frac{1-\pi(X_1)}{\pi(X_1)m(X_1)}\\
&\times \frac{1-p(X_2)}{m(X_2)}\Bigg).
\end{aligned}
$$

The Taylor expansions and change of variable yields

$$
\begin{aligned}
&E\left(B_{hi}(x)(1-\xi_i)\nu_i^{+,m}\right)\\
&= \frac{1}{n^2 g_2} K(0)\frac{1-\pi(x)}{\pi(x)}\frac{1-p(x)}{m(x)} + \frac{1}{2}\frac{h^2}{n^2 g_2} K(0)\frac{1}{m(x)}\left[\frac{1-\pi(x)}{\pi(x)}(1-p(x))\right]'' d_K\\
&- \frac{1}{2}\frac{g_1^2}{n^2 g_2} K(0)\frac{1-\pi(x)}{\pi(x)}\frac{p''(x)}{m(x)} d_K + O\left(\frac{1}{n^2 g_2}\left(h^4 + g_1^2 h^2 + g_1^4 h^{-1}\right)\right).
\end{aligned}
\tag{B.19}
$$

Likewise, for case (b) the following arguments are considered

$$
\begin{aligned}
&E\left(B_{hi}(x)(1-\xi_i)\nu_i^{+,m}\right)\\
&= \frac{1}{n^2 g_1 g_2} \sum_{\substack{j=1\\ j\neq i}}^{n} \sum_{\substack{k=1\\ k\neq i\\ k\neq j}}^{n} E\Bigg(\frac{(\pi(X_i)\pi(X_k))^{-1}}{m(X_i)m(X_k)} K\left(\frac{X_i-X_k}{g_1}\right) K\left(\frac{X_j-X_k}{g_2}\right)\\
&\times B_{hi}(x)(1-\xi_i)\xi_j\xi_k\nu_j\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
&= 2\frac{1}{n^3 g_1 g_2 h}\frac{1}{m(x)}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{k=1\\ k\neq i\\ k\neq j}}^{n} E\bigg(\frac{(\pi(X_i)\pi(X_k))^{-1}}{m(X_i)m(X_k)}K\left(\frac{X_i-X_k}{g_1}\right)K\left(\frac{X_j-X_k}{g_2}\right)\\
&\quad\times K\left(\frac{x-X_i}{h}\right)(1-\xi_i)\xi_j\xi_k\nu_j\bigg)\\
&\quad-\frac{1}{n^4 g_1 g_2 h^2}\frac{1}{m^2(x)}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{k=1\\ k\neq i\\ k\neq j}}^{n}\sum_{l=1}^{n} E\left(\frac{(\pi(X_i)\pi(X_k))^{-1}}{m(X_i)m(X_k)}\right.\\
&\quad\times K\left(\frac{X_i-X_k}{g_1}\right)K\left(\frac{X_j-X_k}{g_2}\right)K\left(\frac{x-X_i}{h}\right)K\left(\frac{x-X_l}{h}\right)(1-\xi_i)\xi_j\xi_k\nu_j\bigg)\\
&\quad+O\left(\frac{1}{n}\left(h^4+\frac{1}{nh}\right)\right)\\
&= E_{b,1}(x)-E_{b,2}(x). \qquad\qquad (B.20)
\end{aligned}
$$

The term $E_{b,1}(x)$ in (B.20) can be verified as

$$
\begin{aligned}
&E_{b,1}(x)\\
&= \frac{1}{n^2 g_1 g_2}\frac{2}{nh}\frac{1}{m(x)}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{k=1\\ k\neq i\\ k\neq j}}^{n} E\bigg[\frac{(\pi(X_i)\pi(X_k))^{-1}}{m(X_i)m(X_k)}K\left(\frac{X_i-X_k}{g_1}\right)K\left(\frac{X_j-X_k}{g_2}\right)\\
&\quad\times K\left(\frac{x-X_i}{h}\right)E\left((1-\xi_i)\xi_j\xi_k\nu_j \mid X_i,X_j,X_k\right)\bigg]\\
&= \frac{(n-1)(n-2)}{n^2 g_1 g_2}\frac{2}{nh}\frac{1}{m(x)}E\left(K\left(\frac{X_1-X_3}{g_1}\right)K\left(\frac{X_2-X_3}{g_2}\right)K\left(\frac{x-X_1}{h}\right)\right.\\
&\quad\times\left.\frac{(\pi(X_1)\pi(X_3))^{-1}}{m(X_1)m(X_3)}(1-\pi(X_1))\pi(X_2)\pi(X_3)(1-p(X_2))\right)\\
&= \frac{(n-1)(n-2)}{n^2 g_1 g_2}\frac{2}{nh}\frac{1}{m(x)}\iiint K\left(\frac{u_1-u_3}{g_1}\right)K\left(\frac{u_2-u_3}{g_2}\right)K\left(\frac{x-u_1}{h}\right)\\
&\quad\times\frac{\pi(u_1)^{-1}}{m(u_1)m(u_3)}(1-\pi(u_1))\pi(u_2)(1-p(u_2))m(u_1)m(u_2)m(u_3)\,du_1du_2du_3\\
&= \frac{(n-1)(n-2)}{n^2 g_1 g_2}\frac{2}{nh}\frac{1}{m(x)}\iiint K\left(\frac{u_1-u_3}{g_1}\right)K\left(\frac{u_2-u_3}{g_2}\right)K\left(\frac{x-u_1}{h}\right)\\
&\quad\times\frac{1-\pi(u_1)}{\pi(u_1)}\pi(u_2)(1-p(u_2))m(u_2)\,du_1du_2du_3.
\end{aligned}
$$

Applying a Taylor expansion and the change of variable $v_2=\frac{u_2-u_3}{g_2}$, $v_3=\frac{u_1-u_3}{g_1}$, $v_1=\frac{x-u_1}{h}$, respectively, it can be shown that

$$
\begin{aligned}
&E_{b,2}(x)\\
&= \frac{(n-1)(n-2)}{n^3 g_1}\frac{2}{h}\frac{1}{m(x)}\iint K\left(\frac{u_1-u_3}{g_1}\right)K\left(\frac{x-u_1}{h}\right)\frac{1-\pi(u_1)}{\pi(u_1)}\pi(u_3)
\end{aligned}
$$

$$
\begin{aligned}
&\times \left(1-p\left(u_3\right)\right) m\left(u_3\right) du_1 du_3 + g_2^2 \frac{(n-1)(n-2)}{n^2 g_1} \frac{1}{nh} \frac{1}{m(x)} d_K \\
&\times \iint K\left(\frac{u_1-u_3}{g_1}\right) K\left(\frac{x-u_1}{h}\right) \frac{1-\pi\left(u_1\right)}{\pi\left(u_1\right)} \left(\pi\left(u_3\right)\left(1-p\left(u_3\right)\right) m\left(u_3\right)\right)'' du_1 du_3 \\
&+ O\left((n-1)(n-2)\, g_2^4 (g_1 h)^{-1} n^{-3}\right) \\
=\ & \frac{(n-1)(n-2)}{n^3 h m(x)} \left\{ 2 \iint K\left(v_3\right) K\left(\frac{x-u_1}{h}\right) \frac{1-\pi\left(u_1\right)}{\pi\left(u_1\right)} \pi\left(u_1-g_1 v_3\right) \right. \\
&\times \left(1-p\left(u_1-g_1 v_3\right)\right) m\left(u_1-g_1 v_3\right) du_1 dv_3 + g_2^2 d_K \\
&\times \iint K\left(v_3\right) K\left(\frac{x-u_1}{h}\right) \frac{1-\pi\left(u_1\right)}{\pi\left(u_1\right)} \Big[\pi\left(u_1-g_1 v_3\right)\left(1-p\left(u_1-g_1 v_3\right)\right) \\
&\times m\left(u_1-g_1 v_3\right) \Big]'' du_1 dv_3 \Big\} + O\left((n-1)(n-2)\, g_2^4 n^{-3} (g_1 h)^{-1}\right) \\
=\ & \frac{(n-1)(n-2)}{n^3 h m(x)} \left\{ 2 \int K\left(\frac{x-u_1}{h}\right) \frac{1-\pi\left(u_1\right)}{\pi\left(u_1\right)} \right. \\
&\times \int K\left(v_3\right) \pi\left(u_1-g_1 v_3\right)\left(1-p\left(u_1-g_1 v_3\right)\right) m\left(u_1-g_1 v_3\right) dv_3 du_1 \\
&+ g_2^2 d_K \int K\left(\frac{x-u_1}{h}\right) \frac{1-\pi\left(u_1\right)}{\pi\left(u_1\right)} \int K\left(v_3\right) \Big[\pi\left(u_1-g_1 v_3\right)\left(1-p\left(u_1-g_1 v_3\right)\right) \\
&\times m\left(u_1-g_1 v_3\right) \Big]'' dv_3 du_1 \Big\} + O\left((n-1)(n-2)\, g_2^4 (g_1 h)^{-1} n^{-3}\right) \\
=\ & \frac{(n-1)(n-2)}{n^3 h m(x)} \left\{ 2 \int K\left(\frac{x-u_1}{h}\right) \left(1-\pi\left(u_1\right)\right)\left(1-p\left(u_1\right)\right) m\left(u_1\right) du_1 \right. \\
&\left. + \left(g_1^2+g_2^2\right) d_K \int K\left(\frac{x-u_1}{h}\right) \frac{1-\pi\left(u_1\right)}{\pi\left(u_1\right)} \left[\pi\left(u_1\right)\left(1-p\left(u_1\right)\right) m\left(u_1\right)\right]'' du_1 \right\} \\
&+ O\left((n-1)(n-2)\, g_1^4 h^{-1} n^{-3}\right) + O\left((n-1)(n-2)\, g_1^2 g_2^2 h^{-1} n^{-3}\right) \\
&+ O\left((n-1)(n-2)\, g_2^4 g_1^{-1} h^{-1} n^{-3}\right) \\
=\ & \frac{(n-1)(n-2)}{n^3} \left\{ 2\left(1-\pi(x)\right)\left(1-p(x)\right) \right. \\
&+ h^2 \frac{1}{m(x)} \left[\left(1-\pi(x)\right)\left(1-p(x)\right) m(x)\right]'' d_K \\
&\left. + \left(g_1^2+g_2^2\right) \frac{1}{m(x)} \frac{1-\pi(x)}{\pi(x)} \left[\pi(x)\left(1-p(x)\right) m(x)\right]'' d_K \right\} \\
&+ O\left((n-1)(n-2)\, h^4 n^{-3}\right) + O\left((n-1)(n-2)\left(g_1^2+g_2^2\right) h^2 n^{-3}\right) \\
&+ O\left((n-1)(n-2)\, g_1^4 h^{-1} n^{-3}\right) + O\left((n-1)(n-2)\, g_1^2 g_2^2 h^{-1} n^{-3}\right) \\
&+ O\left((n-1)(n-2)\, g_2^4 g_1^{-1} h^{-1} n^{-3}\right).
\end{aligned}
$$

Turning to $E_{b,2}(x)$ in (B.20), it can be shown that

$$\begin{aligned}
E_{b,2}^{2}(x) = {} & \frac{1}{n^{4}g_{1}g_{2}h^{2}}\frac{1}{m^{2}(x)}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{k=1\\ k\neq i\\ k\neq j}}^{n}\sum_{l=1}^{n}E\Bigg[\frac{(\pi(X_i)\pi(X_k))^{-1}}{m(X_i)m(X_k)}K\left(\frac{X_i-X_k}{g_1}\right)\\
& \times K\left(\frac{X_j-X_k}{g_2}\right)K\left(\frac{x-X_i}{h}\right)K\left(\frac{x-X_l}{h}\right)\\
& \times(1-\pi(X_i))\pi(X_j)\pi(X_k)(1-p(X_j))\Bigg]\\
= {} & \frac{n(n-1)(n-2)}{n^{3}g_{1}g_{2}h^{2}}\frac{1}{m^{2}(x)}\\
& \times E\left(K\left(\frac{X_1-X_3}{g_1}\right)K\left(\frac{X_2-X_3}{g_2}\right)K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_4}{h}\right)\right.\\
& \left.\times\ \frac{1-\pi(X_1)}{m(X_1)\pi(X_1)}\pi(X_2)(1-p(X_2))\frac{1}{m(X_3)}\right)\\
= {} & \frac{(n-1)(n-2)}{n^{3}g_{1}g_{2}h^{2}m^{2}(x)}\iiiint K\left(\frac{u_1-u_3}{g_1}\right)K\left(\frac{u_2-u_3}{g_2}\right)K\left(\frac{x-u_1}{h}\right)K\left(\frac{x-u_4}{h}\right)\\
& \left.\times\ \frac{1-\pi(u_1)}{\pi(u_1)}\pi(u_2)(1-p(u_2))m(u_2)m(u_4)\,du_1du_2du_3du_4\right)
\end{aligned}$$

By following similar arguments as in the proof of $E_{b,1}(x)$, it suffices to show

$$\begin{aligned}
& E_{b,2}(x)\\
= {} & \frac{(n-1)(n-2)}{n^{3}}\Bigg\{(1-\pi(x))(1-p(x))\\
& +\frac{h^{2}}{2}\left[\frac{((1-\pi)(1-p)m)''(x)}{m(x)}+\frac{m''(x)}{m(x)}(1-\pi(x))(1-p(x))\right]d_K\\
& +\frac{(g_1^{2}+g_2^{2})}{2}\frac{1-\pi(x)}{\pi(x)}\frac{(\pi(1-p)m)''(x)}{m(x)}d_K\Bigg\}\\
& +O\left((n-1)(n-2)n^{-3}\left(g_2^{4}+h^{4}+h^{2}g_2^{2}+h^{2}g_2^{2}+g_1^{2}g_2^{2}\right)\right).
\end{aligned}$$

Combining $E_{b,1}(x)$ and $E_{b,2}(x)$, one obtain

$$\begin{aligned}
& E\left(B_{hi}(x)(1-\xi_i)\nu_i^{+,m}\right)\\
= {} & \frac{(n-1)(n-2)}{n^{3}}\Bigg\{(1-\pi(x))(1-p(x))+\frac{h^{2}}{2m(x)}d_K\left[(1-\pi(x))(1-p(x))m(x)\right]''\\
& -m''(x)(1-\pi(x))(1-p(x))+\frac{(g_1^{2}+g_2^{2})}{2m(x)}\frac{1-\pi(x)}{\pi(x)}d_K\left[\pi(x)(1-p(x))m(x)\right]''\Bigg\}\\
& +O\left((n-1)(n-2)n^{-3}\left(g_2^{4}+h^{4}+g_1^{2}h^{2}+g_2^{2}h^{2}+g_1^{2}g_2^{2}\right)\right). \qquad \text{(B.21)}
\end{aligned}$$

Taking into account (B.19) and (B.21), then the second expectation in (B.15) is

$$\begin{aligned}
&E_2(x)\\
&= \frac{(n-1)(n-2)}{n^2}\Bigg((1-\pi(x))(1-p(x)) + \frac{h^2}{2m(x)} d_K \{ [(1-\pi(x))(1-p(x))m(x)]''\\
&\quad - m''(x)(1-\pi(x))(1-p(x)) \} + \frac{(g_1^2+g_2^2)}{2m(x)} d_K \frac{1-\pi(x)}{\pi(x)} [\pi(x)(1-p(x))m(x)]'' \Bigg)\\
&\quad + \frac{1}{ng_2} K(0) \frac{1-\pi(x)}{\pi(x)} \frac{1-p(x)}{m(x)}\\
&\quad + O\left(n^{-2}(n-1)(n-2)\left(h^2+g_1^2+g_2^2\right)^2 \right) + O\left(n^{-1} g_2^{-1}\left(h^2+g_1^2\right)\right). \qquad \text{(B.22)}
\end{aligned}$$

The derivation of the dominant terms of the bias of $1-\widehat{p}_h^{\text{MI-NW}}(x)$ is complete after joining together (B.17) and (B.22) and the fact that $h \to 0, g_1 \to 0, g_2 \to 0, nh \to \infty, ng_1 \to \infty$ and $ng_2 \to \infty$ as $n \to \infty$.

Conditioning on the observed data (O) and the resampling data in Step 1. (R) of Section 3.3.2, the variance of $1-\widehat{p}_h^{\text{MI-NW}}(x)$ is

$$\begin{aligned}
\sigma^2_{h,\text{MI-NW}}(x) &= E\left(\operatorname{Var}\left(1-\widehat{p}_h^{\text{MI-NW}}(x) \mid O,R\right)\right)\\
&\quad + \operatorname{Var}\left(E\left(1-\widehat{p}_h^{\text{MI-NW}}(x) \mid O,R\right)\right) = V_1 + V_2. \qquad \text{(B.23)}
\end{aligned}$$

To prove (B.23), the auxiliary results in Lemmas D.9 – D.11 are needed.

The first term in (B.23) is

$$\begin{aligned}
V_1 &= \frac{1}{M} E\left[\operatorname{Var}\left(\sum_{i=1}^{n} \frac{1}{nh}\frac{1}{m(x)} K\left(\frac{x-X_i}{h}\right)\left(\xi_i\nu_i + (1-\xi_i)\nu_i^{+,1}\right) \mid O,R \right)\right]\\
&= \frac{1}{M} E\left[\frac{1}{nh^2}\frac{1}{m^2(x)} K^2\left(\frac{x-X_1}{h}\right)(1-\xi_1) E\left(\nu_1^{+,1} \mid O,R\right)\right]\\
&\quad - \frac{1}{M} E\left[\frac{1}{nh^2}\frac{1}{m^2(x)} K^2\left(\frac{x-X_1}{h}\right)(1-\xi_1) E\left(\nu_1^{+,1} \mid O,R\right)^2\right]\\
&= V_{11} - V_{12}.
\end{aligned}$$

After Lemmas D.9 and D.11, respectively, we obtain

$$\begin{aligned}
V_{11} &= \frac{c_K}{Mnh}\frac{(1-\pi(x))(1-p(x))}{m(x)}\left(1+O\left(g_1^2+g_2^2+h^2\right)\right)\\
V_{12} &= \frac{c_K}{Mnh}\frac{(1-\pi(x))(1-p(x))^2}{m(x)}\left(1+O\left(g_1^2+g_2^2+h^2\right)\right).
\end{aligned}$$

Thus,

$$V_1 = \frac{c_K}{Mnh} \frac{(1-\pi(x))\, p(x)\, (1-p(x))}{m(x)} \left(1 + O\left(g_1^2 + g_2^2 + h^2\right)\right). \tag{B.24}$$

Next, turning to the second term of (B.23):

$$\begin{aligned} V_2 &= \operatorname{Var}\left(E\left[E\left(1-\widehat{p}_h^{\text{MI-NW}}(x)|O,R\right) \mid O\right]\right) &(B.25)\\ &\quad + E\left[\operatorname{Var}\left(E\left(1-\widehat{p}_h^{\text{MI-NW}}(x) \mid O,R\right) \mid O\right)\right] \\ &= V_{21} + V_{22}. &(B.26) \end{aligned}$$

The first term in (B.26) is

$$\begin{aligned} V_{21} &= \frac{1}{n^2h^2}\frac{1}{m^2(x)}\operatorname{Var}\left(\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\left(\xi_i\nu_i + (1-\xi_i)\, E\left(\nu_1^{+,1} \mid O,R\right)\right)\right) \\ &= \frac{1}{nh^2}\frac{1}{m^2(x)}\operatorname{Var}\left(K\left(\frac{x-X_1}{h}\right)\xi_1\nu_1\right) \\ &\quad + \frac{1}{nh^2}\frac{1}{m^2(x)}\operatorname{Var}\left(K\left(\frac{x-X_1}{h}\right)(1-\xi_1)\, E\left(\nu_1^{+,1} \mid O,R\right)\right) \\ &\quad + 2\frac{1}{nh^2}\frac{1}{m^2(x)}\operatorname{Cov}\left(K\left(\frac{x-X_1}{h}\right)\xi_1\nu_1, K\left(\frac{x-X_1}{h}\right)(1-\xi_1)\, E\left(\nu_1^{+,1} \mid O,R\right)\right) \\ &\quad + \frac{1}{n^2h^2}\frac{1}{m^2(x)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\operatorname{Cov}\left[K\left(\frac{x-X_i}{h}\right)\left(\xi_i\nu_i + (1-\xi_i)\, E\left(\nu_i^{+,1} \mid O,R\right)\right)\right. \\ &\quad \left. \times\, K\left(\frac{x-X_j}{h}\right)\left(\xi_j\nu_j + (1-\xi_j)\, E\left(\nu_j^{+,1} \mid O,R\right)\right)\right] \\ &= I_1 + I_2 + I_3 + I_4. \end{aligned}$$

It can be shown that

$$I_1 = \frac{c_K}{nh}\frac{\pi(x)\,(1-p(x))}{m(x)} + O\left(n^{-1}\right), \tag{B.27}$$

$$I_2 = \frac{c_K}{nh}\frac{(1-\pi(x))\,(1-p(x))^2}{m(x)} + O\left(n^{-1}\right), \tag{B.28}$$

$$I_3 = -2\frac{1}{n}\frac{\pi(x)(1-\pi(x))(1-p(x))^2}{m(x)} + O\left(n^{-1}h^2\right) = O\left(n^{-1}\right). \tag{B.29}$$

The term I_4 is simplified to

$$I_4 = (1-p(x))^2(\pi^2(x)-1) + 2\frac{n-1}{nh^2}\frac{1}{m^2(x)}I_{41} + \frac{n-1}{nh^2}\frac{1}{m^2(x)}I_{42} + O(h^2), \tag{B.30}$$

where

$$\begin{aligned}
I_{41} &= E\left[K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_2}{h}\right)\xi_1\nu_1\left(1-\xi_2\right)E\left(\nu_2^{+,1}|O,R\right)\right],\\
I_{42} &= E\Bigg[K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_2}{h}\right)\left(1-\xi_1\right)\left(1-\xi_2\right)E\left(\nu_1^{+,1}|O,R\right)\\
&\quad\times E\left(\nu_2^{+,1}|O,R\right)\Bigg].
\end{aligned}$$

The first term I_{41} can be written as

$$\begin{aligned}
I_{41} &= E\left[K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_2}{h}\right)\xi_1\nu_1\left(1-\xi_2\right)E\left(\nu_2^{+,1}|O,R\right)\right]\\
&\simeq \frac{1}{n^2g_1g_2}\sum_{i=1}^n\sum_{j=1}^n E\Bigg[\xi_1\left(1-\xi_2\right)\xi_i\xi_j\frac{\pi^{-1}\left(X_i\right)m^{-1}\left(X_i\right)}{\pi\left(X_1\right)m\left(X_1\right)}\\
&\quad\times K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_2}{h}\right)K\left(\frac{X_2-X_i}{g_2}\right)K\left(\frac{X_i-X_j}{g_1}\right)\nu_1\nu_j\Bigg].
\end{aligned}$$

There are different cases to be considered. Note that the cases $i=2$ and $j=2$ give $I_{41}=0$.

Case (a1): $i=j=1$

$$\begin{aligned}
I_{41} &\simeq \frac{1}{n^2g_1g_2}E\Bigg[\xi_1\left(1-\xi_2\right)\frac{\pi^{-1}\left(X_1\right)m^{-1}\left(X_1\right)}{\pi\left(X_1\right)m\left(X_1\right)}\\
&\quad\times K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_2}{h}\right)K\left(\frac{X_2-X_1}{g_2}\right)K\left(0\right)\nu_1\Bigg]\\
&= O\left(\frac{1}{n^2g_1g_2}hg_2\right)=O\left(\frac{h}{n}\frac{1}{ng_1}\right)=o\left(n^{-1}h\right).
\end{aligned}$$

Case (a2): $i=1, j\neq 1,2$

$$\begin{aligned}
I_{41} &\simeq \frac{n-2}{n^2g_1g_2}E\Bigg[\xi_1\left(1-\xi_2\right)\xi_3\frac{\pi^{-1}\left(X_1\right)m^{-1}\left(X_1\right)}{\pi\left(X_1\right)m\left(X_1\right)}\\
&\quad\times K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_2}{h}\right)K\left(\frac{X_2-X_1}{g_2}\right)K\left(\frac{X_1-X_3}{g_1}\right)\nu_1\nu_3\Bigg]\\
&= \frac{n-2}{n^2g_1g_2}\iiint\pi\left(x_1\right)\left(1-\pi\left(x_2\right)\right)\pi\left(x_3\right)\frac{\pi^{-1}\left(x_1\right)m^{-1}\left(x_1\right)}{\pi\left(x_1\right)m\left(x_1\right)}K\left(\frac{x-x_1}{h}\right)K\left(\frac{x-x_2}{h}\right)\\
&\quad\times K\left(\frac{x_2-x_1}{g_2}\right)K\left(\frac{x_1-x_3}{g_1}\right)\left(1-p\left(x_1\right)\right)\left(1-p\left(x_3\right)\right)m\left(x_1\right)m\left(x_2\right)m\left(x_3\right)dx_1dx_2dx_3\Bigg]\\
&= \frac{n-2}{n^2g_1g_2}\iiint\left(1-\pi\left(x_2\right)\right)\pi\left(x_3\right)\frac{m^{-1}\left(x_1\right)}{\pi\left(x_1\right)}K\left(\frac{x-x_1}{h}\right)K\left(\frac{x-x_2}{h}\right)\\
&\quad\times K\left(\frac{x_2-x_1}{g_2}\right)K\left(\frac{x_1-x_3}{g_1}\right)\left(1-p\left(x_1\right)\right)\left(1-p\left(x_3\right)\right)m\left(x_2\right)m\left(x_3\right)dx_1dx_2dx_3\Bigg].
\end{aligned}$$

Applying 3 changes of variables and Taylor expansion, we obtain

$$\begin{aligned}I_{41} &\simeq \frac{n-2}{n^2 g_1 g_2} g_1 g_2 h \iiint \left(1-\pi\left(x-hu_1+g_2u_2\right)\right)\pi\left(x-hu_1-g_1u_3\right)\frac{m^{-1}\left(x-hu_1\right)}{\pi\left(x-hu_1\right)}\\ &\quad \times\left(1-p\left(x-hu_1\right)\right)\left(1-p\left(x-hu_1-g_1u_3\right)\right)m\left(x-hu_1+g_2u_2\right)m\left(x-hu_1-g_1u_3\right)\\ &\quad \times K\left(u_1\right)K\left(u_1-\frac{g_2}{h}u_2\right)K\left(u_2\right)K\left(u_3\right)du_1du_2du_3.\end{aligned}$$

According to Lemma D.12, the term I_{41} follows:

- If $\frac{g_2}{h}\to 0$,

$$\begin{aligned}I_{41} &\simeq \frac{n-2}{n^2}h\left(\left(1-\pi\left(x\right)\right)\left(1-p\left(x\right)\right)^2 m\left(x\right)+O\left(h+g_1+g_2\right)\right)\\ &\quad\times\left(c_K+O\left(\frac{g_2^2}{h^2}\right)\right)\\ &= n^{-1}h\left(1-\pi\left(x\right)\right)\left(1-p\left(x\right)\right)^2 m\left(x\right)c_K+O\left(n^{-1}h^{-1}g_2^2\right)\\ &\quad+O\left(n^{-1}h\left(h+g_1+g_2\right)\right)+O\left(\frac{1}{n^2}h\right)\\ &= n^{-1}h\left(1-\pi\left(x\right)\right)\left(1-p\left(x\right)\right)^2 m\left(x\right)c_K+O\left(n^{-1}h\right).\end{aligned}$$

- If $\frac{g_2}{h}\to C$,

$$I_{41}\simeq n^{-1}h\left(1-\pi\left(x\right)\right)\left(1-p\left(x\right)\right)^2 m\left(x\right)c_{K,C}+O\left(nh^{-1}\right).$$

- If $\frac{g_2}{h}\to\infty$,

$$\begin{aligned}I_{41} &\simeq \frac{n-2}{n^2}h\left(\left(1-\pi\left(x\right)\right)\left(1-p\left(x\right)\right)^2 m\left(x\right)+O\left(h+g_1+g_2\right)\right)O\left(\frac{h}{g_2}\right)\\ &= O\left(\frac{h}{n}\frac{h}{g_2}\right)=o\left(\frac{h}{n}\right).\end{aligned}$$

Case (a3) $i\neq 1,2, j=1$

$$\begin{aligned}I_{41} &\simeq \frac{n-2}{n^2 g_1 g_2}E\Bigg[\xi_1\left(1-\xi_2\right)\xi_3\frac{\pi^{-1}\left(X_3\right)m^{-1}\left(X_3\right)}{\pi\left(X_1\right)m\left(X_1\right)}\\ &\quad K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_2}{h}\right)K\left(\frac{X_2-X_3}{g_2}\right)K\left(\frac{X_3-X_1}{g_1}\right)\nu_1\Bigg]\\ &= \frac{n-2}{n^2 g_1 g_2}\iiint \pi\left(x_1\right)\left(1-\pi\left(x_2\right)\right)\pi\left(x_3\right)\frac{\pi^{-1}\left(x_3\right)m^{-1}\left(x_3\right)}{\pi\left(x_1\right)m\left(x_1\right)}K\left(\frac{x-x_1}{h}\right)K\left(\frac{x-x_2}{h}\right)\\ &\quad K\left(\frac{x_2-x_3}{g_2}\right)K\left(\frac{x_3-x_1}{g_1}\right)\left(1-p\left(x_1\right)\right)m\left(x_1\right)m\left(x_2\right)m\left(x_3\right)dx_1dx_2dx_3\end{aligned}$$

$$= \frac{n-2}{n^2 g_1 g_2} \iiint (1-\pi(x_2))(1-p(x_1))\, m(x_2)$$
$$K\left(\frac{x-x_1}{h}\right) K\left(\frac{x-x_2}{h}\right) K\left(\frac{x_2-x_3}{g_2}\right) K\left(\frac{x_3-x_1}{g_1}\right) dx_1 dx_2 dx_3.$$

Applying changes of variables and Taylor expansion, it is evidently

$$\begin{aligned} I_{41} &\simeq \frac{n-2}{n^2 g_1 g_2} g_1 g_2 h \iiint (1-\pi(x-hu_1+g_1u_3+g_2u_2))(1-p(x-hu_1)) \\ &\times m(x-hu_1+g_1u_3+g_2u_2)\, K(u_1)\, K\left(u_1-\frac{g_1}{h}u_3-\frac{g_2}{h}u_2\right) \\ &\times K(u_2)\, K(u_3)\, du_1 du_2 du_3. \end{aligned}$$

Following Lemma D.13

- If $\frac{g_1}{h} \to 0$ and $\frac{g_2}{h} \to 0$ then

$$\begin{aligned} I_{41} &\simeq \frac{n-2}{n^2} h \iiint (1-\pi(x-hu_1+g_1u_3+g_2u_2))(1-p(x-hu_1)) \\ &\times m(x-hu_1+g_1u_3+g_2u_2) \\ &K(u_1)\, K\left(u_1-\frac{g_1}{h}u_3-\frac{g_2}{h}u_2\right) K(u_2)\, K(u_3)\, du_1 du_2 du_3 \\ &= n^{-1}h(1-\pi(x))(1-p(x))\, m(x)\, c_K + o\left(n^{-1}h\right). \end{aligned}$$

- If one pilot bandwidth verifies $\frac{g_i}{h} \to 0$ and the other one $\frac{g_j}{h} \to C$ then

$$I_{41} = n^{-1}h(1-\pi(x))(1-p(x))\, m(x)\, c_{K,C} + o\left(n^{-1}h\right).$$

  where $c_{K,C} = \iint K(u_1)\, K(u_2)\, K(u_1 - Cu_2)\, du_1 du_2$.

- If $\frac{g_1}{h} \to C_1$ and $\frac{g_2}{h} \to C_2$ then

$$I_{41} = n^{-1}h(1-\pi(x))(1-p(x))\, m(x)\, c_{K,C_1,C_2} + o\left(n^{-1}h\right).$$

  where $c_{K,C_1,C_2} = \iint K(u_1)\, K(u_2)\, K(u_1 - C_1u_3 - C_2u_2)\, du_1 du_2$

- If $\frac{g_1}{h} \to \infty$ or $\frac{g_2}{h} \to \infty$ then

$$I_{41} = o\left(n^{-1}h\right).$$

Case (a4): $i, j \neq 1, 2, i = j$

$$I_{41} \simeq \frac{n-2}{n^2 g_1 g_2} E\left[\xi_1 (1-\xi_2)\, \xi_{3j} \frac{\pi^{-1}(X_3)\, m^{-1}(X_3)}{\pi(X_1)\, m(X_1)}\right.$$

$$
\begin{aligned}
&K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_2}{h}\right)K\left(\frac{X_2-X_3}{g_2}\right)K\left(0\right)\nu_1\nu_3\Big]\\
&=\frac{n-2}{n^2g_1g_2}K\left(0\right)\iiint\pi\left(x_1\right)\left(1-\pi\left(x_2\right)\right)\pi\left(x_3\right)\frac{\pi^{-1}\left(x_3\right)m^{-1}\left(x_3\right)}{\pi\left(x_1\right)m\left(x_1\right)}K\left(\frac{x-x_1}{h}\right)\\
&K\left(\frac{x-x_2}{h}\right)K\left(\frac{x_2-x_3}{g_2}\right)\left(1-p\left(x_1\right)\right)\left(1-p\left(x_3\right)\right)m\left(x_1\right)m\left(x_2\right)m\left(x_3\right)dx_1dx_2dx_3\\
&=\frac{n-2}{n^2g_1g_2}K\left(0\right)\iiint\left(1-p\left(x_1\right)\right)\left(1-\pi\left(x_2\right)\right)m\left(x_2\right)\left(1-p\left(x_3\right)\right)\\
&K\left(\frac{x-x_1}{h}\right)K\left(\frac{x-x_2}{h}\right)K\left(\frac{x_2-x_3}{g_2}\right)dx_1dx_2dx_3.
\end{aligned}
$$

Again, after 3 changes of variable and applying Taylor expansions,

$$
\begin{aligned}
I_{41}&=\frac{n-2}{n^2g_1g_2}h^2g_2K\left(0\right)\iiint\left(1-p\left(x-hu_1\right)\right)\left(1-\pi\left(x-hu_2\right)\right)m\left(x-hu_2\right)\\
&\quad\left(1-p\left(x-hu_2-g_2u_3\right)\right)K\left(u_1\right)K\left(u_1\right)K\left(u_3\right)du_1du_2dx_3\\
&=\frac{1}{ng_1}h^2K\left(0\right)\left(1-p\left(x\right)\right)^2\left(1-\pi\left(x\right)\right)m\left(x\right)+O\left(\frac{h^2}{n^2g_1}\right)\\
&=\frac{1}{ng_1}h^2K\left(0\right)\left(1-p\left(x\right)\right)^2\left(1-\pi\left(x\right)\right)m\left(x\right)+O\left(\frac{h^2}{ng_1}\right).
\end{aligned}
$$

Case (a5): $i,j\neq1,2,i\neq j$

$$
\begin{aligned}
I_{41}&\simeq\frac{\left(n-2\right)\left(n-3\right)}{n^2g_1g_2}E\left[\xi_1\left(1-\xi_2\right)\xi_3\xi_4\frac{\pi^{-1}\left(X_3\right)m^{-1}\left(X_3\right)}{\pi\left(X_1\right)m\left(X_1\right)}\right.\\
&\quad\left.K\left(\frac{x-X_1}{h}\right)K\left(\frac{x-X_2}{h}\right)K\left(\frac{X_2-X_3}{g_2}\right)K\left(\frac{X_3-X_4}{g_1}\right)\nu_1\nu_4\right]\\
&=\frac{\left(n-2\right)\left(n-3\right)}{n^2g_1g_2}\iiiint\pi\left(x_1\right)\left(1-\pi\left(x_2\right)\right)\pi\left(x_3\right)\pi\left(x_4\right)\frac{\pi^{-1}\left(x_3\right)m^{-1}\left(x_3\right)}{\pi\left(x_1\right)m\left(x_1\right)}\\
&\quad\times K\left(\frac{x-x_1}{h}\right)K\left(\frac{x-x_2}{h}\right)K\left(\frac{x_2-x_3}{g_2}\right)K\left(\frac{x_3-x_4}{g_1}\right)\left(1-p\left(x_1\right)\right)\\
&\quad\times\left(1-p\left(x_4\right)\right)m\left(x_1\right)m\left(x_2\right)m\left(x_3\right)m\left(x_4\right)dx_1dx_2dx_3dx_4\\
&=\frac{\left(n-2\right)\left(n-3\right)}{n^2g_1g_2}\iiiint\left(1-\pi\left(x_2\right)\right)\pi\left(x_4\right)K\left(\frac{x-x_1}{h}\right)K\left(\frac{x-x_2}{h}\right)\\
&\quad\times K\left(\frac{x_2-x_3}{g_2}\right)K\left(\frac{x_3-x_4}{g_1}\right)\left(1-p\left(x_1\right)\right)\left(1-p\left(x_4\right)\right)\\
&\quad\times m\left(x_2\right)m\left(x_4\right)dx_1dx_2dx_3dx_4\\
&=\frac{\left(n-2\right)\left(n-3\right)}{n^2g_1g_2}\int\left(1-p\left(x_1\right)\right)K\left(\frac{x-x_1}{h}\right)dx_1\\
&\quad\times\iiiint\left(1-\pi\left(x_2\right)\right)\pi\left(x_4\right)K\left(\frac{x-x_2}{h}\right)K\left(\frac{x_2-x_3}{g_2}\right)K\left(\frac{x_3-x_4}{g_1}\right)\\
&\quad\times\left(1-p\left(x_4\right)\right)m\left(x_2\right)m\left(x_4\right)dx_2dx_3dx_4.
\end{aligned}
$$

We can perform 3 changes of variables

$$\begin{aligned}
I_{41} &= \frac{(n-2)(n-3)}{n^2 g_1 g_2} h \left[(1-p(x)) + O\left(h^2\right) \right] h g_1 g_2 \iiiint (1-\pi(x-hu_2)) \\
&\quad \times \pi(x-hu_2-g_2u_3-g_1u_4)(1-p(x-hu_2-g_2u_3-g_1u_4)) \\
&\quad \times m(x-hu_2) m(x-hu_2-g_2u_3-g_1u_4) K(u_2) K(u_3) K(u_4) du_2 dx_3 dx_4 \\
&= \frac{(n-2)(n-3)}{n^2} h^2 \left[(1-p(x)) + O\left(h^2\right) \right] \left[(1-\pi(x)) \pi(x) (1-p(x)) m^2(x) \right. \\
&\quad \left. + O\left(h^2+g_1^2+g_2^2\right) \right] \\
&= h^2 (1-p(x))^2 (1-\pi(x)) \pi(x) m^2(x) - \frac{5}{n} h^2 (1-p(x))^2 (1-\pi(x)) \\
&\quad \times \pi(x) m^2(x) + O\left(\frac{h^2}{n^2}\right) + O\left(h^2\left(h^2+g_1^2+g_2^2\right)\right).
\end{aligned}$$

Collecting Cases (a1) − (a5), the first term in I_4 has got different expressions depending on the bandwidths h, g_1 and g_2

- If $\frac{g_1}{h} \to 0$ and $\frac{g_2}{h} \to 0$ then

$$\begin{aligned}
I_{41} &= h^2 (1-p(x))^2 (1-\pi(x)) \pi(x) m^2(x) - \frac{5}{n} h^2 (1-p(x))^2 (1-\pi(x)) \\
&\quad \times \pi(x) m^2(x) + \frac{1}{n} h (1-\pi(x)) (1-p(x))^2 m(x) c_K \\
&\quad + \frac{1}{n} h (1-\pi(x)) (1-p(x)) m(x) c_K \\
&\quad + \frac{1}{ng_1} h^2 K(0) (1-p(x))^2 (1-\pi(x)) m(x) \\
&\quad + o\left(\frac{h^2}{ng_1}\right) + O\left(\frac{h^2}{n^2}\right) + O\left(h^2\left(h^2+g_1^2+g_2^2\right)\right) + o\left(\frac{h}{n}\right).
\end{aligned}$$

- If $\frac{g_1}{h} \to 0$ and $\frac{g_2}{h} \to C$ then

$$\begin{aligned}
I_{41} &= h^2 (1-p(x))^2 (1-\pi(x)) \pi(x) m^2(x) - \frac{5}{n} h^2 (1-p(x))^2 (1-\pi(x)) \\
&\quad \times \pi(x) m^2(x) + \frac{1}{n} h (1-\pi(x)) (1-p(x))^2 m(x) c_{K,C} \\
&\quad + \frac{1}{n} h (1-\pi(x)) (1-p(x)) m(x) c_{K,C} \\
&\quad + \frac{1}{ng_1} h^2 K(0) (1-p(x))^2 (1-\pi(x)) m(x) \\
&\quad + o\left(h^2 \frac{1}{ng_1}\right) + O\left(\frac{h^2}{n^2}\right) + O\left(h^2\left(h^2+g_1^2+g_2^2\right)\right) + o\left(\frac{h}{n}\right)
\end{aligned}$$

where $c_{K,C} = \iint K(u_1) K(u_2) K(u_1 - Cu_2) du_1 du_2$.

- If $\frac{g_1}{h} \to C$ and $\frac{g_2}{h} \to 0$ then

$$\begin{aligned}
I_{41} &= h^2 (1-p(x))^2 (1-\pi(x)) \pi(x) m^2(x) - \frac{5}{n} h^2 (1-p(x))^2 (1-\pi(x)) \\
&(x) m^2(x) + \frac{1}{n} h (1-\pi(x)) (1-p(x))^2 m(x) c_K \\
&+ \frac{1}{n} h (1-\pi(x)) (1-p(x)) m(x) c_{K,C} \\
&+ \frac{1}{ng_1} h^2 K(0) (1-p(x))^2 (1-\pi(x)) m(x) \\
&+ o\left(h^2 \frac{1}{ng_1}\right) + O\left(\frac{h^2}{n^2}\right) + O\left(h^2\left(h^2 + g_1^2 + g_2^2\right)\right) + o\left(\frac{h}{n}\right)
\end{aligned}$$

where $c_{K,C} = \iint K(u_1) K(u_2) K(u_1 - Cu_2) du_1 du_2$.

- If $\frac{g_1}{h} \to C_1$ and $\frac{g_2}{h} \to C_2$ then

$$\begin{aligned}
I_{41} &= h^2 (1-p(x))^2 (1-\pi(x)) \pi(x) m^2(x) - \frac{5}{n} h^2 (1-p(x))^2 (1-\pi(x)) \\
&(x) m^2(x) + \frac{1}{n} h (1-\pi(x)) (1-p(x))^2 m(x) c_{K,C_2} \\
&+ \frac{1}{n} h (1-\pi(x)) (1-p(x)) m(x) c_{K,C_1,C_2} \\
&+ \frac{1}{ng_1} h^2 K(0) (1-p(x))^2 (1-\pi(x)) m(x) + o\left(h^2 \frac{1}{ng_1}\right) \\
&+ O\left(\frac{h^2}{n^2}\right) + O\left(h^2\left(h^2 + g_1^2 + g_2^2\right)\right) + o\left(\frac{h}{n}\right)
\end{aligned}$$

where $c_{K,C_1,C_2} = \iint K(u_1) K(u_2) K(u_1 - C_1 u_3 - C_2 u_2) du_1 du_2$ and $c_{K,C} = \iint K(u_1) K(u_2) K(u_1 - Cu_2) du_1 du_2$.

- If any pilot bandwidth verifies $\frac{g_i}{h} \to \infty$ then

$$\begin{aligned}
I_{41} &= h^2 (1-p(x))^2 (1-\pi(x)) \pi(x) m^2(x) - \frac{5}{n} h^2 (1-p(x))^2 (1-\pi(x)) \\
&\times \pi(x) m^2(x) + \frac{1}{ng_1} h^2 K(0) (1-p(x))^2 (1-\pi(x)) m(x) + o\left(h^2 \frac{1}{ng_1}\right) \\
&+ O\left(\frac{h^2}{n^2}\right) + O\left(h^2\left(h^2 + g_1^2 + g_2^2\right)\right) + o\left(\frac{h}{n}\right).
\end{aligned}$$

To summarize, if $g_1/h \to C_1 \geq 0$ and $g_2/h \to C_2 \geq 0$, then

$$\begin{aligned}
&2\frac{n-1}{nh^2}\frac{1}{m^2(x)}I_{41} \\
&= 2\pi(x)(1-\pi(x))(1-p(x))^2 \\
&\quad + \frac{2}{nh}\left[\frac{(1-\pi(x))(1-p(x))^2}{m(x)}c_{K,C_2} + \frac{(1-\pi(x))(1-p(x))}{m(x)}c_{K,C_1,C_2}\right] \\
&\quad + \frac{2}{ng_1}\frac{(1-\pi(x))(1-p(x))^2}{m(x)}K(0) + O\left(n^{-1}h^{-1}\right) + O\left(n^{-1}g_1^{-1}\right),
\end{aligned} \tag{B.31}$$

where $c_{K,C_1,C_2} = \iiint K(u)K(v)K(w)K(u+C_1v+C_2w)dudvdw$ and $c_{K,C_2} = \iint K(u)K(v)K(u+C_2v)dudv$. Note that if $C_1 = C_2 = 0$ then $c_{K,C_1,C_2} = c_{K,C_2} = c_K = \int K^2(v)dv$. On the other hand, if $C_1 = \infty$ then $c_{K,C_1,C_2} = 0$, and $C_1 = \infty$ implies $c_{K,C_1,C_2} = c_{K,C_2} = 0$.
The result for I_{42} can be proved in similar manner as above:

$$\begin{aligned}
&\frac{n-1}{nh^2}\frac{1}{m^2(x)}I_{42} \\
&= (1-\pi(x))^2(1-p(x))^2 \\
&\quad + \frac{1}{nh}\Big[\frac{(1-\pi(x))^2(1-p(x))^2}{\pi(x)m(x)}(c_{K,C_1,C_2} + 2d_{K,C_1,C_2}) \\
&\quad + \frac{(1-\pi(x))^2(1-p(x))}{\pi(x)m(x)}d_{K,C_1,C_2}\Big] \\
&\quad + \frac{2}{ng_1}\frac{(1-\pi(x))^2(1-p(x))^2}{\pi(x)m(x)}K(0) + O\left(n^{-1}h^{-1}\right) + O\left(n^{-1}g_1^{-1}\right),
\end{aligned} \tag{B.32}$$

where $d_{K,C_1,C_2} = \iiint K(u)K(v)K(w)K(u+C_1v+C_2(u+w))dudvdw$. Again, if $C_1 = C_2 = 0$ then $d_{K,C_1,C_2} = c_K$, whereas if $C_1 = \infty$ or $C_2 = \infty$ then $d_{K,C_1,C_2} = 0$.
Finally, rejoining (B.27) – (B.32), we arrive to the following expression for V_{21}:

$$\begin{aligned}
V_{21} &= \frac{1}{nh}\frac{1-p(x)}{m(x)}\left\{\pi(x)c_K + (1-\pi(x))\left[c_{K,C_1,C_2} + \frac{1-\pi(x)}{\pi(x)}d_{K,C_1,C_2}\right.\right. \\
&\quad \left.\left. + \ (1-p(x))\left(c_K + 2c_{K,C_2} + \frac{1-\pi(x)}{\pi(x)}(c_{K,C_1,C_2} + 2d_{K,C_1,C_2})\right)\right]\right\} \\
&\quad + \frac{2}{ng_1}\frac{(1-\pi(x))(1-p(x))^2}{\pi(x)m(x)}K(0) + O\left(n^{-1}h^{-1}\right) + O\left(n^{-1}g_1^{-1}\right).
\end{aligned} \tag{B.33}$$

The term V_{22} may also be checked in a similar manner as V_{21}:

$$V_{22} = \frac{1}{Mnh}\frac{c_K}{m^2(x)}\frac{(1-\pi(x))^2}{\pi(x)}p(x)(1-p(x)) + O\left(M^{-1}n^{-1}h^{-1}\right). \tag{B.34}$$

Note that adding V_1 in (B.24), V_{21} in (B.33) and V_{22} in (B.34) gives (3.21). $\square$

Appendix C

Proofs of the results in Chapter 4

Proof of Theorem 4.1

Theorem 4.1 (Asymptotic representation). Suppose that Assumptions $1-9$ hold, then, for $x \in I$ and $t \in [a, b]$ such that $\widehat{S}^c_{h_2}(t \mid x) > 1 - \widehat{p}^c_{h_1}(x)$, an iid representation for $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ is

$$\widehat{S}^c_{0,h_1,h_2}(t \mid x) - S_0(t \mid x) = \sum_{i=1}^{n} \eta_{h_1,h_2}(T_i, \delta_i, \xi_i, \nu_i, t, x) + R_n(t, x)$$

where

$$\begin{aligned}\eta_{h_1,h_2}(T_i, \delta_i, \xi_i, \nu_i, t, x) = &-\frac{S(t \mid x)}{p(x)}\widetilde{B}_{h_2 i}(x)\zeta(T_i, \delta_i, \xi_i, \nu_i, t, x) \\ &-\frac{(1-p(x))(1-S(t \mid x))}{p^2(x)}\widetilde{B}_{h_1 i}(x)\zeta(T_i, \delta_i, \xi_i, \nu_i, \tau_0, x)\end{aligned}$$

with $\zeta(T_i, \delta_i, \xi_i, \nu_i, t, x)$ in (2.21),

$$\widetilde{B}_{h_j i}(x) = \frac{1}{m(x)}\frac{1}{nh_j}K\left(\frac{x - X_i}{h_j}\right), \text{ for } \ j = 1, 2,$$

and $R_n(t, x)$ can be shown to satisfy

$$\sup_{a \le t \le b, x \in I} \mid R_n(t, x) \mid = O\left((nh)^{-3/4}(\log n)^{3/4}\right) \text{ a.s.}$$

Proof. Departing from (4.2), adding and subtracting suitable terms, the difference $\widehat{S}^c_{0,h_1,h_2}(t \mid x) - S_0(t \mid x)$ can be written as:

$$\widehat{S}^c_{0,h_1,h_2}(t \mid x) - S_0(t \mid x)$$

$$
\begin{aligned}
=&\frac{\widehat{S}^c_{h_2}(t \mid x)-\left(1-\widehat{p}^c_{h_1}(x)\right)}{\widehat{p}^c_{h_1}(x)}-\frac{S(t \mid x)-(1-p(x))}{p(x)} \pm \frac{\widehat{S}^c_{h_2}(t \mid x)-\left(1-\widehat{p}^c_{h_1}(x)\right)}{p(x)}\\
&\pm \frac{S(t \mid x)-(1-p(x))}{p(x)} \pm \frac{S(t \mid x)-(1-p(x))}{\widehat{p}^c_{h_1}(x)}\\
&\pm \frac{\widehat{p}^c_{h_1}(x)-p(x)}{p^2(x)}\left[S(t \mid x)-(1-p(x))\right]\\
=&I+II+III+R_1(t,x)+R_2(t,x)+R_3(t,x),
\end{aligned}
$$

where

$$
\begin{aligned}
I =&\frac{\widehat{S}^c_{h_2}(t \mid x)-S(t \mid x)}{p(x)}, \quad II=\frac{\widehat{p}^c_{h_1}(x)-p(x)}{p(x)},\\
III =&-\frac{\left[S(t \mid x)-(1-p(x))\right]\left(\widehat{p}^c_{h_1}(x)-p(x)\right)}{p^2(x)}\\
R_{n1}(t,x) =&\frac{\left(\widehat{S}^c_{h_2}(t \mid x)-S(t \mid x)\right)\left(p(x)-\widehat{p}^c_{h_1}(x)\right)}{\widehat{p}^c_{h_1}(x)p(x)},\\
R_{n2}(t,x) =&-\frac{\left(\widehat{p}^c_{h_1}(x)-p(x)\right)^2}{\widehat{p}^c_{h_1}(x)p(x)} \quad \text{and}\\
R_{n3}(t,x) =&\frac{\left[S(t \mid x)-(1-p(x))\right]\left(\widehat{p}^c_{h_1}(x)-p(x)\right)^2}{\widehat{p}^c_{h_1}(x)p^2(x)}.
\end{aligned}
$$

Note that the dominant term in (4.5) is derived from $I+II+III$ since the terms $R_{n1}(t,x)+R_{n2}(t,x)+R_{n3}(t,x)$ are negligible. From Theorem 2.2, the next iid representation of I is obtained:

$$
I=\frac{-S(t \mid x)}{p(x)}\sum_{i=1}^n \widetilde{B}_{h_2 i}(x)\,\zeta\left(T_i,\delta_i,\xi_i,\nu_i,t,x\right)+O\left((nh_2)^{-3/4}(\log n)^{3/4}\right) \text{ a.s.} \qquad \text{(C.1)}
$$

In sequel, the sum of the terms $II+III$ is studied,

$$
II+III=\frac{1-S(t \mid x)}{p^2(x)}\left(\widehat{p}^c_{h_1}(x)-p(x)\right).
$$

From Theorem 3.1, the next iid representation is obtained:

$$
\begin{aligned}
II+III=&-\frac{(1-S(t \mid x))(1-p(x))}{p^2(x)}\sum_{i=1}^n \widetilde{B}_{h_1 i}(x)\,\zeta\left(T_i,\delta_i,\xi_i,\nu_i,\tau_0,x\right)\\
&+O\left((nh_1)^{-3/4}(\log n)^{3/4}\right) \text{ a.s.} \qquad \text{(C.2)}
\end{aligned}
$$

From (C.1) and (C.2) we get the dominant terms of the iid representation in (4.5). In order to show that the term $R_{n1}(t,x)$ is negligible, the results from Corollary 2.1 and Corollary 3.1 are applied to obtain

$$\sup_{a\le t\le b, x\in I} |R_{1n}(t,x)| = O\left(n^{-1}(h_1h_2)^{-1/2}\log n\right) \text{ a.s.}$$

Similarly, the other remainders terms are studied:

$$R_{2n}(t,x) + R_{3n}(t,x) = \frac{S(t\mid x)-1}{p^2(x)} \frac{\left(\widehat{p}^c_{h_1}(x)-p(x)\right)^2}{\widehat{p}^c_{h_1}(x)}.$$

Again, using Corollary 3.1,

$$\sup_{a\le t\le b, x\in I} |R_{2n}(t,x)+R_{3n}(t,x)| = O\left((nh_1)^{-1}\log n\right) \text{ a.s.}$$

This concludes the proof.

□

Proof of Proposition 4.1

Proposition 4.1 Suppose that Assumptions $1-9$ are satisfied, then, the asymptotic bias and variance of the dominant term in the iid representation of $\widehat{S}^c_{0,h_1,h_2}(t\mid x)$ are, respectively,

$$\mu^c_{h_1,h_2}(t,x) = h_1^2 B_{c,1}(t,x) + h_2^2 B_{c,2}(t,x) + O\left(h_1^4\right) + O\left(h_2^4\right),$$

and

$$\sigma^2_{c,h_1,h_2}(t,x) = \frac{1}{nh_1}s^2_{c,1}(x) + \frac{1}{nh_2}\left(s^2_{c,2}(t,x) + 2s^2_{c,3}(t,x)\right) + O(n^{-1}h_2) + O((nh_2)^{-1}h_1).$$

Notice the dominant terms in the bias are

$$\begin{aligned} B_{c,1}(t,x) =& \frac{1-S(t\mid x)\,d_K}{2p^2(x)\,m(x)}\left[2\,(1-p(x))'\,m'(x) + (1-p(x))''m(x)\right] \\ &- \frac{(1-p(x))\,(1-S(t\mid x))\,d_K}{p^2(x)} \int_0^{\tau_0} \frac{G'(v^-\mid x)}{1-G(v^-\mid x)} \frac{d}{ds}\left(\frac{S'(s\mid x)}{S(s\mid x)}\right)\Bigg|_{s=v^-} dv \end{aligned}$$

and

$$\begin{aligned} B_{c,2}(t,x) =& \frac{d_K}{2p(x)\,m(x)}\left(2S'(t^-\mid x)\,m'(x) + S''(t^-\mid x)\,m(x)\right) \\ &- \frac{S(t\mid x)\,d_K}{p(x)} \int_0^{t} \frac{G'(v^-\mid x)}{1-G(v^-\mid x)} \frac{d}{ds}\left(\frac{S'(s\mid x)}{S(s\mid x)}\right)\Bigg|_{s=v^-} dv. \end{aligned}$$

Here, $d_K = \int v^2 K(v)dv$ and $S'(t \mid x), S''(t \mid x)$, $(1-p(x))'$, $(1-p(x))''$ refer to the derivatives with respect to x.

The dominant terms in the variance are

$$\begin{aligned}
s^2_{c,1}(t,x) =& \frac{(1-p(x))^2\,(1-S(t\mid x))^2}{p^4(x)m(x)} \int_0^{\tau_0} \frac{dH^1(v\mid x)c_K}{(1-H(v\mid x)+H^{11}(v\mid x))^2},\\
s^2_{c,2}(t,x) =& \frac{S^2(t\mid x)}{p^2(x)m(x)} \int_0^{t} \frac{dH^1(v\mid x)c_K}{(1-H(v\mid x)+H^{11}(v\mid x))^2},\\
s^2_{c,3}(t,x) =& \frac{(1-p(x))\,(1-S(t\mid x))\,S(t\mid x)}{p^3(x)m(x)} \int_0^{t} \frac{dH^1(v\mid x)}{(1-H(v\mid x)+H^{11}(v\mid x))^2}\\
&\times \int K(v)K(v\frac{h_1}{h_2})dv
\end{aligned}$$

where $c_K = \int K^2(v)dv$.

Proof. From Theorem 4.1, the bias of $\widehat{S}^c_{0,h_1,h_2}(t \mid x)$ is asymptotically equal to the expectation of

$$\begin{aligned}
&\sum_{i=1}^n \eta_{h_1,h_2}(T_i,\delta_i,\xi_i,\nu_i,t,x)\\
=&-\frac{1}{nh_2}\frac{S(t\mid x)}{m(x)p(x)}\sum_{i=1}^n K\left(\frac{x-X_i}{h_2}\right)\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x)\\
&-\frac{1}{nh_1}\frac{(1-p(x))(1-S(t\mid x))}{m(x)p^2(x)}\sum_{i=1}^n K\left(\frac{x-X_i}{h_1}\right)\zeta(T_i,\delta_i,\xi_i,\nu_i,\tau_0,x)\\
=&I_S+I_p+II_S+II_p, \qquad \text{(C.3)}
\end{aligned}$$

where

$$\begin{aligned}
I_S =& -\frac{1}{nh_2}\frac{S(t\mid x)}{m(x)p(x)}\left[\sum_{i=1}^n K\left(\frac{x-X_i}{h_2}\right)\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x)\right.\\
&\left.-E\left(\sum_{i=1}^n K\left(\frac{x-X_i}{h_2}\right)\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x)\right)\right], \qquad \text{(C.4)}\\
I_p =& -\frac{1}{nh_1}\frac{(1-p(x))(1-S(t\mid x))}{m(x)p^2(x)}\left[\sum_{i=1}^n K\left(\frac{x-X_i}{h_1}\right)\zeta(T_i,\delta_i,\xi_i,\nu_i,\tau_0,x)\right.\\
&\left.-E\left(\sum_{i=1}^n K\left(\frac{x-X_i}{h_1}\right)\zeta(T_i,\delta_i,\xi_i,\nu_i,\tau_0,x)\right)\right], \qquad \text{(C.5)}\\
II_S =& -\frac{1}{nh_2}\frac{S(t\mid x)}{m(x)p(x)}E\left(\sum_{i=1}^n K\left(\frac{x-X_i}{h_2}\right)\zeta(T_i,\delta_i,\xi_i,\nu_i,t,x)\right), \qquad \text{(C.6)}
\end{aligned}$$

$$II_p = -\frac{1}{nh_1}\frac{(1-p(x))(1-S(t\mid x))}{m(x)p^2(x)}E\left(\sum_{i=1}^{n}K\left(\frac{x-X_i}{h_1}\right)\zeta\left(T_i,\delta_i,\xi_i,\nu_i,\tau_0,x\right)\right). \tag{C.7}$$

Note that $E(I_S) = E(I_p) = 0$, thus the asymptotic bias of the estimator $\widehat{S}^c_{0,h_1,h_2}(t\mid x)$ is $II_p + II_S$. From (B.6) and (A.16), it can be shown that, respectively,

$$\begin{aligned}II_p =& -\frac{h_1^2}{2}\frac{(1-p(x))(1-S(t\mid x))}{m(x)p^2(x)}\left(2\Phi'_c\left(x,\tau_0,x\right)m'\left(x\right)+\Phi''_c\left(x,\tau_0,x\right)m\left(x\right)\right)d_K\\ &+O(h_1^4),\\ II_S =& -\frac{h_2^2}{2}\frac{S(t\mid x)}{m(x)p(x)}\left(2\Phi'_c\left(x,t,x\right)m'\left(x\right)+\Phi''_c\left(x,t,x\right)m\left(x\right)\right)d_K+O(h_2^4),\end{aligned} \tag{C.8}$$

where $d_K = \int v^2K(v)dv$, $\Phi'_c(y,t,x)$ and $\Phi''_c(y,t,x)$ are the first and second derivatives with respect to y of $\Phi_c(y,t,x) = E(\zeta(T,\delta,\xi,\nu,t,x)\mid X=y)$. The expression of the bias of $\widehat{S}^c_{0,h_1,h_2}(t\mid x)$ is obtained by plugging-in (D.6), (D.11), (B.8), and (B.9) in (C.8). Recalling (C.4) – (C.7), the asymptotic variance of $\widehat{S}^c_{0,h_1,h_2}(t\mid x)$ is

$$\operatorname{Var}(I_S + I_p + II_S + II_p) = \operatorname{Var}(I_S) + \operatorname{Var}(I_p) + 2\operatorname{Cov}(I_S, I_p). \tag{C.9}$$

Note that

$$\operatorname{Var}\left(I_S\right) = \frac{S^2(t\mid x)}{p^2(x)m^2(x)}(V_1 - V_2),$$

where

$$\begin{aligned}V_1 =& \frac{1}{nh_2^2}E\left(K^2\left(\frac{x-X}{h_2}\right)\zeta^2\left(T,\delta,\xi,\nu,t,x\right)\right),\\ V_2 =& \frac{1}{nh_2^2}\left[E\left(K\left(\frac{x-X}{h_2}\right)\zeta\left(T,\delta,\xi,\nu,t,x\right)\right)\right]^2.\end{aligned}$$

From the results for V_1 and V_2 in (A.19) and (A.18) in Appendix A, one has

$$\operatorname{Var}\left(I_S\right) = \frac{1}{nh_2}\frac{S^2(t\mid x)}{p^2(x)m(x)}c_K\int_0^t\frac{dH^1(v\mid x)}{\left(1-H(v\mid x)+H^{11}(v\mid x)\right)^2}+O(n^{-1}h_2), \tag{C.10}$$

where $c_K = \int K^2(v)dv$. Following the lines of the proof of $\operatorname{Var}(I_S)$ (see also the proof of the asymptotic variance of $1-\widehat{p}^c_{h_1}(t\mid x)$ in Appendix B),

$$\begin{aligned}\operatorname{Var}\left(I_p\right) =& \frac{1}{nh_1}\frac{(1-S(t\mid x))^2(1-p(x))^2}{p^4(x)m(x)}c_K\int_0^{\tau_0}\frac{dH^1(v\mid x)}{\left(1-H(v\mid x)+H^{11}(v\mid x)\right)^2}\\ &+O(n^{-1}h_1).\end{aligned} \tag{C.11}$$

As for the third term in (C.9), one obtains

$$\operatorname{Cov}\left(I_S, I_p\right) = \operatorname{Cov}\left(\widetilde{I}_S, \widetilde{I}_p\right),$$

where

$$\begin{aligned}
\widetilde{I}_S &= -\frac{1}{nh_2}\frac{S(t \mid x)}{m(x)p(x)}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_2}\right)\zeta\left(T_i, \delta_i, \xi_i, \nu_i, t, x\right), \\
\widetilde{I}_p &= -\frac{1}{nh_1}\frac{\left(1-p(x)\right)\left(1-S(t \mid x)\right)}{m(x)p^2(x)}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_1}\right)\zeta\left(T_i, \delta_i, \xi_i, \nu_i, \tau_0, x\right).
\end{aligned}$$

The covariance of $\widetilde{I}_S$ and $\widetilde{I}_p$ is

$$\begin{aligned}
&\operatorname{Cov}\left(\widetilde{I}_S, \widetilde{I}_p\right) \\
=&\frac{1}{n^2h_1h_2}\frac{S(t \mid x)\left(1-p(x)\right)\left(1-S(t \mid x)\right)}{m^2(x)p^3(x)}\operatorname{Cov}\Bigg(\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_2}\right)\zeta\left(T_i, \delta_i, \xi_i, \nu_i, t, x\right), \\
&\sum_{j=1}^{n} K\left(\frac{x-X_j}{h_1}\right)\zeta\left(T_j, \delta_j, \xi_j, \nu_j, \tau_0, x\right)\Bigg) \\
=&\frac{1}{n^2h_1h_2}\frac{S(t \mid x)\left(1-p(x)\right)\left(1-S(t \mid x)\right)}{m^2(x)p^3(x)} \\
&\times\left(\sum_{i=1}^{n}\operatorname{Cov}\left(K\left(\frac{x-X_i}{h_1}\right)\zeta\left(T_i, \delta_i, \xi_i, \nu_i, t, x\right), K\left(\frac{x-X_i}{h_2}\right)\zeta\left(T_i, \delta_i, \xi_i, \nu_i, \tau_0, x\right)\right)\right. \\
&\left.+\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}\operatorname{Cov}\Bigg(K\left(\frac{x-X_i}{h_1}\right)\zeta\left(T_i, \delta_i, \xi_i, \nu_i, t, x\right), K\left(\frac{x-X_j}{h_2}\right)\zeta\left(T_j, \delta_j, \xi_j, \nu_j, \tau_0, x\right)\Bigg)\right) \\
=&\frac{1}{nh_1h_2}\frac{S(t \mid x)\left(1-p(x)\right)\left(1-S(t \mid x)\right)}{m^2(x)p^3(x)}\operatorname{Cov}\Bigg(K\left(\frac{x-X_1}{h_1}\right)\zeta\left(T_1, \delta_1, \xi_1, \nu_1, t, x\right), \\
&K\left(\frac{x-X_1}{h_2}\right)\zeta\left(T_1, \delta_1, \xi_1, \nu_1, \tau_0, x\right)\Bigg).
\end{aligned} \tag{C.12}$$

The covariance term in (C.12) can be worked out as follows:

$$\begin{aligned}
&\operatorname{Cov}\left(K\left(\frac{x-X_1}{h_2}\right)\zeta\left(T_1, \delta_1, \xi_1, \nu_1, t, x\right), K\left(\frac{x-X_1}{h_1}\right)\zeta\left(T_1, \delta_1, \xi_1, \nu_1, \tau_0, x\right)\right) \\
=&E\left(K\left(\frac{x-X_1}{h_1}\right)K\left(\frac{x-X_1}{h_2}\right)\zeta\left(T_1, \delta_1, \xi_1, \nu_1, t, x\right)\zeta\left(T_1, \delta_1, \xi_1, \nu_1, \tau_0, x\right)\right) \\
&-E\left(K\left(\frac{x-X_1}{h_1}\right)\zeta\left(T_1, \delta_1, \xi_1, \nu_1, t, x\right)\right)E\left(K\left(\frac{x-X_1}{h_2}\right)\zeta\left(T_1, \delta_1, \xi_1, \nu_1, \tau_0, x\right)\right) \\
=&\gamma-\alpha\beta.
\end{aligned} \tag{C.13}$$

By applying a change of variable, Taylor expansion and following the lines of Lemma 6 of López-Cheda et al. (2017b), one attains

$$\gamma = h_1 \int_0^t \frac{dH^1(v \mid x)}{\left(1 - H(v \mid x) + H^{11}(v \mid x)\right)^2} m(x) \int K(v) K\left(\frac{h_1}{h_2} v\right) dv + O(h_1^2). \tag{C.14}$$

From Lemma D.1, the terms α and β are

$$\begin{aligned}
\alpha =& \frac{h_1^3}{2m(x)} \left(2\Phi_c'\left(x, \tau_0, x\right) m'\left(x\right) + \Phi_c''\left(x, \tau_0, x\right) m\left(x\right)\right) d_K + O(h_1^5),\\
\beta =& \frac{h_2^3}{2m(x)} \left(2\Phi_c'\left(x, t, x\right) m'\left(x\right) + \Phi_c''\left(x, t, x\right) m\left(x\right)\right) d_K + O(h_2^5).
\end{aligned} \tag{C.15}$$

Substituting (C.13) – (C.15) in (C.12) yield

$$\begin{aligned}
&\operatorname{Cov}(\widetilde{I}_S, \widetilde{I}_p)\\
=& \frac{1}{nh_1h_2} \frac{S(t \mid x)\left(1 - p(x)\right)\left(1 - S(t \mid x)\right)}{m^2(x) p^3(x)} \left[h_1 \Phi_2^c(x, t, x) m(x) \int K(v) K\left(\frac{h_1}{h_2} v\right) dv \right.\\
&- \frac{h_1^3 h_2^3}{4m^2(x)} d_K^2 \left(2\Phi_c'\left(x, t, x\right) m'\left(x\right) + \Phi_c''\left(x, t, x\right) m\left(x\right)\right)\\
&\left. \times \left(2\Phi_c'\left(x, \tau_0, x\right) m'\left(x\right) + \Phi_c''\left(x, \tau_0, x\right) m\left(x\right)\right) + O(h_1^5 h_2^5) + O(h_1^2) \right]\\
=& 2\frac{1}{nh_2} \frac{S(t \mid x)\left(1 - p(x)\right)\left(1 - S(t \mid x)\right)}{m(x) p^3(x)} \int_0^t \frac{dH^1(v \mid x)}{\left(1 - H(v \mid x) + H^{11}(v \mid x)\right)^2}\\
&\times \int K(v) K\left(\frac{h_1}{h_2} v\right) dv + O((nh_2)^{-1} h_1).
\end{aligned} \tag{C.16}$$

Substituting (C.10), (C.11) and (C.16) into (C.9) completes the proof of (4.8). □

Proof of Theorem 4.2

Theorem 4.2 Suppose that Assumptions 1 – 9 and 11 are satisfied, then, for $x \in I$ and $t \in [a, b]$ such that $\widehat{S}_{h_2}^c(t \mid x) > 1 - \widehat{p}_{h_1}^c(x)$, it follows that

(i) If $nh_1^5 \to 0$ and $nh_2^5 \to 0$, then

$$\sqrt{\frac{nh_1h_2}{h_1 + h_2}} \left(\widehat{S}_{0,h_1,h_2}^c\left(t \mid x\right) - S_0\left(t \mid x\right)\right) \xrightarrow{d} N(0, s_c^2(t, x)),$$

where

$$s_c^2(t,x) = \begin{cases} s_{c,1}^2(t,x) & \text{if } \dfrac{h_1}{h_2} \to 0 \\[2ex] s_{c,2}^2(t,x) & \text{if } \dfrac{h_2}{h_1} \to 0 \\[2ex] \dfrac{C_2 s_{c,1}^2(t,x)}{C_1+C_2} + \dfrac{C_1\left(s_{c,2}^2(t,x) + 2s_{c,3}^2(t,x)\right)}{C_1+C_2} & \text{if } \dfrac{h_1}{h_2} \to \dfrac{C_1}{C_2} \end{cases}$$

with $s_{c,1}^2(t,x)$, $s_{c,2}^2(t,x)$, $s_{c,3}^2(t,x)$ given in (4.11) – (4.13), respectively, and C_1 and C_2 are constants.

(ii) If $nh_1^5 \to 0$ and $nh_2^5 \to C_2^5 > 0$, then

$$\sqrt{\frac{nh_1h_2}{h_1+h_2}}\left(\widehat{S}_{0,h_1,h_2}^c(t \mid x) - S_0(t \mid x)\right) \xrightarrow{d} N(0, s_{c,1}^2(t,x)).$$

(iii) If $nh_1^5 \to C_1^5 > 0$ and $nh_2^5 \to 0$, then

$$\sqrt{\frac{nh_1h_2}{h_1+h_2}}\left(\widehat{S}_{0,h_1,h_2}^c(t \mid x) - S_0(t \mid x)\right) \xrightarrow{d} N(0, s_{c,2}^2(t,x)).$$

(iv) If $nh_1^5 \to C_1^5 > 0$ and $nh_2^5 \to C_2^5 > 0$, then

$$\sqrt{\frac{nh_1h_2}{h_1+h_2}}\left(\widehat{S}_{0,h_1,h_2}^c(t \mid x) - S_0(t \mid x)\right) \xrightarrow{d} N(B_c(t,x), s_c^2(t,x))$$

where

$$B_c(t,x) = \sqrt{\frac{C_1C_2}{C_1+C_2}}\left(C_1^2 B_{c,1}(t,x) + C_2^2 B_{c,2}(t,x)\right),$$

with $B_{c,1}(t,x)$ and $B_{c,2}(t,x)$ defined in (4.9) and (4.10), and $s_c^2(t,x)$ is given in (4.14).

Proof. We must characterize the asymptotic distribution of $\sqrt{\frac{nh_1h_2}{h_1+h_2}}\left(\widehat{S}_{0,h_1,h_2}^c(t \mid x) - S_0(t \mid x)\right)$. By Theorem 4.1 and following the proof of Proposition 4.1, this is the same asymptotic distribution of

$$\sqrt{\frac{nh_1h_2}{h_1+h_2}}\sum_{i=1}^{n}\eta_{h_1,h_2}(T_i,\delta_i,\xi_i,\nu_i,t,x) = \sqrt{\frac{nh_1h_2}{h_1+h_2}}(I_S + I_p + II_S + II_p), \tag{C.17}$$

with I_S, I_p, II_S and II_p are given in (C.4) - (C.8) since, by Assumption 11.,

$$\sqrt{\frac{nh_1h_2}{h_1+h_2}}O\left((nh_1)^{-3/4}(\log n)^{3/4}+(nh_2)^{-3/4}(\log n)^{3/4}\right)=o(1).$$

From (C.8), it is obvious that the expectation of (C.17) is

$$\sqrt{\frac{nh_1h_2}{h_1+h_2}}(II_p+II_S)=\sqrt{\frac{nh_1h_2}{h_1+h_2}}\left(h_1^2B_{c,1}(t,x)+h_2^2B_{c,2}(t,x)+O\left(h_1^4\right)+O\left(h_2^4\right)\right), \tag{C.18}$$

where $B_{c,1}(t,x)$ and $B_{c,2}(t,x)$ are given in (4.9) and (4.10), respectively.
Following (C.9) – (C.11) and (C.16), we obtain that the variance of (C.17) is

$$\begin{aligned}\sigma^2_{c,h_1,h_2}(t,x)=&\frac{h_2}{h_1+h_2}s^2_{c,1}(t,x)+\frac{h_1}{h_1+h_2}\left(s^2_{c,2}(t,x)+2s^2_{c,3}(t,x)\right)\\&+O\left(\frac{h_1h_2^2}{h_1+h_2}\right)+O\left(\frac{h_1^2h_2}{h_1+h_2}\right)\\&<s^2_{c,1}(t,x)+s^2_{c,2}(t,x)+2s^2_{c,3}(t,x)+o(1)<\infty,\end{aligned} \tag{C.19}$$

whereby $h_i>0$ and $h_i/(h_1+h_2)<1$, for $i=1,2$. The terms $s^2_{c,1}(t,x)$, $s^2_{c,2}(t,x)$ and $s^2_{c,3}(t,x)$ given in (4.11) – (4.13) are finite as a consequence $\sigma^2_{c,h_1,h_2}(t,x)<\infty$. By applying Lindeberg's theorem for triangular arrays (Billingsley, 1968), we conclude the asymptotic normality of $\widehat{S}^c_{0,h_1,h_2}(t\mid x)$.
Next, we provide the expressions of the bias and variance according to the rate at which the bandwidths h_1,h_2 tend to zero.

(i) If $nh_i^5\frac{h_j}{h_1+h_2}\to 0$, for $i,j=1,2$ $i\neq j$, then the bias is negligible, applying (C.18),

$$\sqrt{\frac{nh_1h_2}{h_1+h_2}}(II_p+II_S)=o(1).$$

As for the expression of the asymptotic variance in (C.19), if $h_1/h_2\to 0$ then

$$\begin{aligned}\sigma^2_{c,h_1,h_2}(t,x)=&\frac{1}{h_1/h_2+1}s^2_{c,1}(t,x)+\frac{h_1/h_2}{h_1/h_2+1}\left(s^2_{c,2}(t,x)+2s^2_{c,3}(t,x)\right)\\&\to s^2_{c,1}(t,x),\end{aligned}$$

and if $h_2/h_1\to 0$ then $\sigma^2_{c,h_1,h_2}(t,x)\to s^2_{c,2}(t,x)$.

Finally, if $h_1/h_2\to C_1/C_2$,

$$\sigma^2_{c,h_1,h_2}(t,x)\to\frac{C_2}{C_1+C_2}s^2_{c,1}(t,x)+\frac{C_1}{C_1+C_2}\Big(s^2_{c,2}(t,x)+2s^2_{c,3}(t,x)\Big). \tag{C.20}$$

(ii) If $nh_1^5 \to 0$ and $nh_2^5 \to C_2^5 > 0$, then $h_2/h_1 \to \infty$, thus,

$$\sqrt{nh_1^5 \frac{1}{h_1/h_2+1}} B_{c,1}(t,x) + \sqrt{C_2^5 \frac{1}{1+h_2/h_1}} B_{c,2}(t,x) = o(1).$$

Under these assumptions, the variance is $\sigma^2_{c,h_1,h_2}(t,x) \to s^2_{c,1}(t,x)$.

(iii) If $nh_1^5 \to C_1^5 > 0$ and $nh_2^5 \to 0$, which implies that $h_2/h_1 \to 0$, then $\sqrt{\frac{nh_1h_2}{h_1+h_2}}(II_p + II_S) \to o(1)$ and the variance verifies $\sigma^2_{c,h_1,h_2}(t,x) \to s^2_{c,2}(t,x)$.

(iv) If $nh_1^5 \to C_1^5$ and $nh_2^5 \to C_2^5$, this implies $h_2/h_1 \to C_2/C_1$, so the bias term is asymptotically

$$\begin{aligned}&\sqrt{C_1^5 \frac{h_2/h_1}{1+h_2/h_1}} B_{c,1}(t,x) + \sqrt{C_2^5 \frac{h_1/h_2}{1+h_1/h_2}} B_{c,2}(t,x)\\ =&\sqrt{\frac{C_1C_2}{C_1+C_2}} \left(C_1^2 B_{c,1}(x) + C_2^2 B_{c,2}(t,x) \right).\end{aligned}$$

Analogously, it can be shown under this assumption that the expression $\sigma^2_{c,h_1,h_2}(t,x)$ is (C.20).

□

Appendix D

Auxiliary Lemmas

Lemma D.1 Let us define $\Phi_c\left(y,t,x\right)=\mathrm{E}\left(\zeta\left(T,\delta,\xi,\nu,t,x\right)\mid X=y\right)$ with $\zeta\left(T,\delta,\xi,\nu,t,x\right)$ given in (2.21). Suppose that $K(v)$ is a kernel function fulfilling Assumption 8, and X is a random variable with density function $m(x)$ fulfilling Assumption 2(i), then,

$$\begin{aligned}&\mathrm{E}\left(K\left(\frac{x-X}{h}\right)\zeta\left(T,\delta,\xi,\nu,t,x\right)\mid X=x\right)\\&=h\Phi_c\left(x,t,x\right)m\left(x\right)+\frac{1}{2}h^3d_K\frac{d^2}{dy^2}\left(\Phi_c\left(y,t,x\right)m\left(y\right)\right)\Big|_{y=x}+O\left(h^5\right),\end{aligned}$$

where $d_K=\int v^2K(v)dv$.

Proof. By conditioning on $X=y$, we have

$$\begin{aligned}&\mathrm{E}\left[K\left(\frac{x-X}{h}\right)\mathrm{E}\left(\zeta\left(T,\delta,\xi,\nu,t,x\right)\mid X=y\right)\right]\\&=\int K\left(\frac{x-y}{h}\right)\mathrm{E}\left(\zeta\left(T,\delta,\xi,\nu,t,x\right)\mid X=y\right)m(y)dy.\end{aligned}\tag{D.1}$$

Let us recall that $\Phi\left(y,t,x\right)=\mathrm{E}\left(\zeta\left(T,\delta,\xi,\nu,t,x\right)\mid X=y\right)$ and denote $\Phi(y,t,x)m(y)\equiv(\Phi m)(y)$ for convenience. After applying a change of variable and a Taylor's expansion (D.1) becomes

$$\begin{aligned}&\int K\left(\frac{x-y}{h}\right)\left(\Phi m\right)(y)\,dy\\&=\int K(z)\left(\Phi m\right)(x-hz)hdz\\&=\int K(z)\Bigg((\Phi m)(x)-hz\frac{d}{dy}\left(\Phi m\right)(y)\mid_{y=x}+\int\frac{1}{2}h^2z^2\frac{d^2}{dy^2}\left(\Phi m\right)(y)\mid_{y=x}\Bigg)hdz+O(h^5)\\&=(\Phi m)(x)h\int K(z)dz-h^2\frac{d}{dy}\left(\Phi m\right)(y)\mid_{y=x}\int zK(z)dz\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}h^3\frac{d^2}{dy^2}\left(\Phi m\right)(y)\mid_{y=x}\int z^2K(z)dz+O(h^5)\\
=\ &\left(\Phi m\right)(x)\,h+\frac{1}{2}h^3\frac{d^2}{dy^2}\left(\left(\Phi m\right)(y)\right)|_{y=x}\int z^2K(z)dz+O\left(h^5\right),
\end{aligned}
$$

and the proof concludes by taking into account the definitions of $\Phi(y,t,x)$, $m(y)$ and d_K. □

Lemma D.2 Let us define $\Phi_c\left(y,t,x\right)=E\left(\zeta\left(T,\delta,\xi,\nu,t,x\right)\mid X=y\right)$, with $\zeta\left(T,\delta,\xi,\nu,t,x\right)$ given in (2.21). Then,

$$
\Phi_c\left(y,t,x\right)=\int_0^t\frac{dH^1\left(v\mid y\right)}{J\left(v^-\mid x\right)}-\int_0^t\frac{J\left(v^-\mid y\right)}{J^2\left(v^-\mid x\right)}dH^1\left(v\mid x\right). \tag{D.2}
$$

Therefore, $\Phi_c\left(x,t,x\right)=0$.

Proof. The conditional expectation of $\zeta\left(T,\delta,\xi,\nu,t,x\right)$ is

$$
\begin{aligned}
\Phi_c\left(y,t,x\right)&=E\left(\frac{\mathbf{1}\left(T\leq t,\delta=1\right)}{J(T^-\mid x)}\mid X=y\right)\\
&\quad-E\left(\int_0^t\left(\mathbf{1}\left(T\geq v\right)+\mathbf{1}\left(T<v,\xi\nu=1\right)\right)\frac{dH^1\left(v\mid x\right)}{J^2\left(v^-\mid x\right)}\mid X=y\right)\\
&=A_1-A_2.
\end{aligned}\tag{D.3}
$$

The first term in (D.3) is

$$
A_1=\int_0^t\frac{\phi\left(v,y\right)dH\left(v\mid y\right)}{J\left(T^-\mid x\right)}=\int_0^t\frac{dH^1\left(v\mid y\right)}{J\left(v^-\mid x\right)}, \tag{D.4}
$$

where $\phi\left(v,y\right)=E\left(\delta\mid T=v,X=y\right)$. As for the second term in (D.3) we have:

$$
\begin{aligned}
A_2&=\int_0^t\Big(E(\mathbf{1}\left(T\geq v\right)\mid X=y)+E(\mathbf{1}\left(T<v,\xi\nu=1\right)\mid X=y)\Big)\frac{dH^1\left(v\mid x\right)}{J^2\left(v^-\mid x\right)}\\
&=\int_0^tJ(v^-\mid y)\frac{dH^1\left(v\mid x\right)}{J^2\left(v^-\mid x\right)}.
\end{aligned}\tag{D.5}
$$

Therefore, substituting (D.4) and (D.5) in (D.3) we obtain (D.2). Then, if we evaluate $\Phi_c\left(y,t,x\right)$ at $y=x$

$$
\Phi_c\left(x,t,x\right)=\int_0^t\frac{dH^1\left(v\mid x\right)}{J\left(v^-\mid x\right)}-\int_0^tJ(v^-\mid x)\frac{dH^1\left(v\mid x\right)}{J^2\left(v^-\mid x\right)}=0.
$$

This concludes the proof. □

Lemma D.3 Let us define $\Phi_c\left(y,t,x\right)=\mathrm{E}\left(\zeta\left(T,\delta,\xi,\nu,t,x\right)\mid X=y\right)$, with $\zeta\left(T,\delta,\xi,\nu,t,x\right)$ given in (2.21), and consider $\Phi\left(y,t,x\right)$ in (2.31). The first derivatives of $\Phi_c\,(y,t,x)$ and $\Phi\,(y,t,x)$ with respect to y evaluated at x verify

$$\Phi_c'\,(x,t,x) = \Phi'\,(x,t,x) = -\frac{S'(t^- \mid x)}{S(t^- \mid x)}, \tag{D.6}$$

where $S'(t \mid x)$ stands for the derivative of $S(t \mid x)$ with respect to x.

Proof. From (2.12) and (2.13) we have $J\,(t \mid x) = (1 - G(t \mid x))\,(1 - F(t \mid x))$ and $dH^1\,(t \mid x) = (1 - G(t \mid x))\,dF(t \mid x)$, respectively. Then, the function $\Phi_c(y,t,x)$ in (D.2) can be written as:

$$\begin{aligned}\Phi_c\,(y,t,x) &= \int_0^t \frac{dH^1\,(v \mid y)}{J\,(v^- \mid x)} - \int_0^t \frac{J\,(v^- \mid y)}{J^2\,(v^- \mid x)} dH^1\,(v \mid x)\\ &= \int_0^t \frac{1 - G\,(v^- \mid y)}{1 - G\,(v^- \mid x)} \left[\frac{dF\,(v \mid y)}{1 - F\,(v^- \mid x)} - \frac{1 - F\,(v^- \mid y)}{(1 - F\,(v^- \mid x))^2} dF\,(v \mid x)\right].\end{aligned}$$

The derivative of $\Phi_c\,(y,t,x)$ with respect to y evaluated at $y = x$ verifies

$$\Phi_c'\,(x,t,x) = \int_0^t \frac{d}{dy}\left[\frac{dF\,(v \mid y)}{1 - F\,(v^- \mid x)} - \frac{1 - F\,(v^- \mid y)}{(1 - F\,(v^- \mid x))^2} dF\,(v \mid x)\right]\Big|_{y=x}. \tag{D.7}$$

Note that $F(t \mid x) = 1 - S(t \mid x) = p(x)\,(1 - S_0(t \mid x))$ and $dF(t \mid x) = -p(x)dS_0(t \mid x)$. Therefore,

$$\begin{aligned}\Phi_c'\,(x,t,x) = &-\int_0^t p'(x)\frac{dS_0(v \mid x)}{S(v^- \mid x)} - \int_0^t p(x)\frac{dS_0'(v \mid x)}{S(v^- \mid x)} - \int_0^t p'(x)p(x)\frac{dS_0(v \mid x)}{S^2(v^- \mid x)}\\ &+ \int_0^t p'(x)S_0(v^- \mid x)\frac{p(x)dS_0(v \mid x)}{S^2(v^- \mid x)} + \int_0^t p(x)S_0'(v^- \mid x)\frac{p(x)dS_0(v \mid x)}{S^2(v^- \mid x)},\end{aligned}$$

where $p'(x)$ and $S_0'(t \mid x)$ stand for the derivatives of $p(x)$ and $S_0(t \mid x)$ with respect to x. Adding and subtracting suitably chosen terms, we have

$$\begin{aligned}&\Phi_c'\,(x,t,x)\\ &= -\int_0^t p'(x)\frac{dS_0(v \mid x)}{S(v^- \mid x)} - \int_0^t p(x)\frac{dS_0'(v \mid x)}{S(v^- \mid x)} - \int_0^t p'(x)p(x)\frac{dS_0(v \mid x)}{S^2(v^- \mid x)}\\ &\quad + \int_0^t p'(x)S_0(v^- \mid x)\frac{p(x)dS_0(v \mid x)}{S^2(v^- \mid x)} + \int_0^t p(x)S_0'(v^- \mid x)\frac{p(x)dS_0(v \mid x)}{S^2(v^- \mid x)}\\ &\quad \pm \int_0^t p'(x)\frac{dS_0(v \mid x)}{S(v^- \mid x)} \pm \int_0^t p(x)p'(x)S_0(v^- \mid x)\frac{dS_0(v \mid x)}{S^2(v^- \mid x)}\\ &= -\int_0^t p'(x)\frac{dS_0(v \mid x)}{S^2(v^- \mid x)} + \int_0^t p'(x)\frac{dS_0(v \mid x)}{S(v^- \mid x)} - \int_0^t p'(x)p(x)S_0(v^- \mid x)\frac{dS_0(v \mid x)}{S^2(v^- \mid x)}\\ &\quad + \int_0^t \frac{(p(x)S_0(v^- \mid x))'\,p(x)dS_0(v \mid x)}{S^2(v^- \mid x)} - \int_0^t \frac{(p(x)dS_0(v \mid x))'}{S(v^- \mid x)}\\ &= I + II,\end{aligned} \tag{D.8}$$

where I and II are the sum of the first three and the last two terms of (D.8), respectively.

Since $dS(t \mid x) = p(x) dS_0(t \mid x)$, then,

$$\begin{aligned} I = & -p(x)p'(x) \int_0^t \frac{dS_0(v \mid x)}{S^2(v^- \mid x)} \\ = & -p'(x) \int_0^t \frac{dS(v \mid x)}{S^2(v^- \mid x)} = p'(x) \left(\frac{1}{S(t^- \mid x)} - \frac{1}{S(0^- \mid x)} \right) = \frac{p'(x)}{S(t^- \mid x)} - p'(x). \end{aligned} \tag{D.9}$$

Similarly for II, we have

$$\begin{aligned} II = & -\int_0^t \frac{d}{ds} \left[\frac{(p(x)S_0(s \mid x))'}{S(s \mid x)} \right] \Big|_{s=v^-} dv \\ = & -\left[\frac{(p(x)S_0(v^- \mid x))'}{S(v^- \mid x)} \right]_0^t = p'(x) - \frac{(p(x)S_0(t^- \mid x))'}{S(t^- \mid x)}. \end{aligned}$$

Taking into account the relationship $(p(x)S_0(t \mid x))' = p'(x) + S'(t^- \mid x)$, then

$$II = p'(x) - \frac{p'(x) + S'(t^- \mid x)}{S(t^- \mid x)}. \tag{D.10}$$

The proof of the result for $\Phi_c(x, t, x)$ concludes by substituting (D.9) and (D.10) in (D.8). As for Beran's estimator, $\Phi(y, t, x)$ in (2.31) can be written as:

$$\begin{aligned} \Phi(y, t, x) &= \int_0^t \frac{dH^1(v \mid y)}{1 - H(v^- \mid x)} - \int_0^t \frac{1 - H(v^- \mid y)}{(1 - H(v^- \mid x))^2} dH^1(v \mid x) \\ &= \int_0^t \frac{1 - G_0(v^- \mid y)}{1 - G_0(v^- \mid x)} \left[\frac{dF(v \mid y)}{1 - F(v^- \mid x)} - \frac{1 - F(v^- \mid y)}{(1 - F(v^- \mid x))^2} dF(v \mid x) \right]. \end{aligned}$$

Thus, using (D.7), the derivative of $\Phi(y, t, x)$ with respect to y evaluated at $y = x$ verifies

$$\Phi'(x, t, x) = \int_0^t \frac{d}{dy} \left[\frac{dF(v \mid y)}{1 - F(v^- \mid x)} - \frac{1 - F(v^- \mid y)}{(1 - F(v^- \mid x))^2} dF(v \mid x) \right] \Big|_{y=x} = \Phi'_c(x, t, x).$$

This completes the proof. □

Lemma D.4 Let us define $\Phi_c(y, t, x) = \mathrm{E}(\zeta(T, \delta, \xi, \nu, t, x) \mid X = y)$, with $\zeta(T, \delta, \xi, \nu, t, x)$ given in (2.21), and consider $\Phi(y, t, x)$ in (2.31). The second derivatives of $\Phi_c(y,t,x)$ and $\Phi(y,t,x)$ with respect to y evaluated at $y=x$ are

$$\Phi_c''(x,t,x) = 2\int_0^t \frac{G'(v^- \mid x)}{1-G(v^- \mid x)} \frac{d}{ds}\left(\frac{S'(s \mid x)}{S(s \mid x)}\right)\Big|_{s=v^-} dv - \frac{S''(t^- \mid x)}{S(t^- \mid x)}, \tag{D.11}$$

where $G(t \mid x) = (1-\pi(x))G_0(t \mid x)$, and

$$\Phi''(x,t,x) = 2\int_0^t \frac{G_0'(v^- \mid x)}{1-G_0(v^- \mid x)} \frac{d}{ds}\left(\frac{S'(s \mid x)}{S(s \mid x)}\right)\Big|_{s=v^-} dv - \frac{S''(t^- \mid x)}{S(t^- \mid x)}.$$

Proof. The second derivative of $\Phi_c(y,t,x)$ in (D.7) evaluated at $y=x$ is

$$\begin{aligned}
\Phi_c''(x,t,x) = 2\int_0^t &\frac{\frac{d}{dy}\left[1-G(v^- \mid y)\right]|_{y=x}}{1-G(v^- \mid x)} \\
&\times \frac{d}{dy}\left[\frac{dF(v \mid y)}{1-F(v^- \mid x)} - \frac{1-F(v^- \mid y)}{(1-F(v^- \mid x))^2} dF(v \mid x)\right]\Big|_{y=x} \\
&+ \int_0^t \frac{d^2}{dy^2}\left[\frac{dF(v \mid y)}{1-F(v^- \mid x)} - \frac{1-F(v^- \mid y)}{(1-F(v^- \mid x))^2} dF(v \mid x)\right]\Big|_{y=x} \\
= I + II.&
\end{aligned}$$

We start by dealing with I. First, note that the second term of the product is

$$\frac{d}{dy}\left[\frac{dF(v \mid y)}{1-F(v^- \mid x)} - \frac{1-F(v^- \mid y)}{(1-F(v^- \mid x))^2} dF(v \mid x)\right]\Big|_{y=x} = -\frac{d}{ds}\left(\frac{S'(s \mid x)}{S(s \mid x)}\right)\Big|_{s=v^-}. \tag{D.12}$$

The first fraction in I is

$$\frac{\frac{d}{dy}\left[1-G(t^- \mid y)\right]\Big|_{y=x}}{1-G(t^- \mid x)} = \frac{S(v^- \mid x)}{J(v^- \mid x)} \frac{d}{dy}\left(\frac{J(v^- \mid y)}{S(v^- \mid y)}\right)\Big|_{y=x} = \frac{J'(v^- \mid x)}{J(v^- \mid x)} - \frac{S'(v^- \mid x)}{S(v^- \mid x)}. \tag{D.13}$$

From (D.12) and (D.13), we can write I as

$$\begin{aligned}
I &= -2\int_0^t \left(\frac{J'(v^- \mid x)}{J(v^- \mid x)} - \frac{S'(v^- \mid x)}{S(v^- \mid x)}\right) \frac{d}{ds}\left(\frac{S'(s \mid x)}{S(s \mid x)}\right)\Big|_{s=v^-} dv \\
&= -2\int_0^t \frac{J'(v^- \mid x)}{J(v^- \mid x)} \frac{d}{ds}\left(\frac{S'(s \mid x)}{S(s \mid x)}\right)\Big|_{s=v^-} dv + \left(\frac{S'(t^- \mid x)}{S(t^- \mid x)}\right)^2.
\end{aligned}$$

From the definition of $J(t \mid x) = (1-G(t \mid x))S(t \mid x)$, it can be shown that

$$\frac{J'(t \mid x)}{J(t \mid x)} = \frac{-G'(t \mid x)S(t \mid x) + (1-G(t \mid x))S'(t \mid x)}{(1-G(t \mid x))S(t \mid x)} = \frac{-G'(t \mid x)}{1-G(t \mid x)} + \frac{S'(t \mid x)}{S(t \mid x)}.$$

As a consequence, the term I can also be written as follows:

$$\begin{aligned}I &= -2\int_0^t \left(\frac{-G'(v^- \mid x)}{1-G(v^- \mid x)} + \frac{S'(v^- \mid x)}{S(v^- \mid x)}\right)\frac{d}{ds}\left(\frac{S'(s \mid x)}{S(s \mid x)}\right)\Big|_{s=v^-} dv + \left(\frac{S'(t^- \mid x)}{S(t^- \mid x)}\right)^2 \\ &= -2\int_0^t \frac{-G'(v^- \mid x)}{1-G(v^- \mid x)}\frac{d}{ds}\left(\frac{S'(s \mid x)}{S(s \mid x)}\right)\Big|_{s=v^-} dv. \qquad \text{(D.14)}\end{aligned}$$

As for II, we have the following result

$$II = \int_0^t \frac{dF''(v \mid x)}{S(v^- \mid x)} - \int_0^t S''(v^- \mid x)\frac{dF(v \mid x)}{S^2(v^- \mid x)}.$$

Using $F(t \mid x) = 1 - S(t \mid x) = p(x) - p(x)S_0(t \mid x)$, then

$$\begin{aligned}II = &-\int_0^t p''(x)\frac{dS_0(v \mid x)}{S^2(v^- \mid x)} - 2p'(x)\frac{dS_0'(v \mid x)}{S(v^- \mid x)} - p(x)\frac{dS_0''(v \mid x)}{S(v^- \mid x)} \\ &+ \int_0^t p(x)\left(2p'(x)S_0'(v^- \mid x) + p(x)S_0''(v^- \mid x)\right)\frac{dS_0(v \mid x)}{S^2(v^- \mid x)},\end{aligned}$$

where $p''(x)$ and $S_0''(t \mid x)$ stand for the second derivatives of $p(x)$ and $S_0(t \mid x)$ with respect to x. Adding and subtracting suitable terms, II can be written as:

$$\begin{aligned}II = &-\int_0^t p''(x)\frac{dS_0(v \mid x)}{S^2(v^- \mid x)} - 2p'(x)\frac{dS_0'(v \mid x)}{S(v^- \mid x)} - p(x)\frac{dS_0''(v \mid x)}{S(v^- \mid x)} \pm p''(x)\frac{dS_0(v \mid x)}{S(v^- \mid x)} \\ &+ \int_0^t p(x)\left(2p'(x)S_0'(v^- \mid x) + p(x)S_0''(v^- \mid x) \pm p''(x)S_0(v^- \mid x)\right)\frac{dS_0(v \mid x)}{S^2(v^- \mid x)} \\ &= A + B, \qquad \text{(D.15)}\end{aligned}$$

where

$$\begin{aligned}A = &-\int_0^t p''(x)\frac{dS_0(v \mid x)}{S^2(v^- \mid x)} + \int_0^t p''(x)\frac{dS_0(v \mid x)}{S(v^- \mid x)} \\ &-\int_0^t p(x)p''(x)S_0(v^- \mid x)\frac{dS_0(v \mid x)}{S^2(v^- \mid x)}, \\ B = &\int_0^t \left(p(x)S_0(v^- \mid x)\right)''\frac{p(x)dS_0(v \mid x)}{S^2(v^- \mid x)} - \int_0^t \frac{1}{S(v^- \mid x)}\left(p(x)dS_0(v \mid x)\right)''.\end{aligned}$$

Recalling that $dS(t \mid x) = p(x)dS_0(t \mid x)$, A simplifies to

$$\begin{aligned}A &= \int_0^t p''(x)\left(-1 + S(v^- \mid x) - p(x)S_0(v^- \mid x)\right)\frac{dS_0(v \mid x)}{S^2(v^- \mid x)} \\ &= p''(x)\int_0^t \frac{p(x)dS_0(v \mid x)}{S^2(v^- \mid x)} = p''(x)\int_0^t \frac{dS(v \mid x)}{S^2(v^- \mid x)} \\ &= -p''(x)\left(-\frac{1}{S(t^- \mid x)} + 1\right) = \frac{p''(x)}{S(t^- \mid x)} - p''(x). \qquad \text{(D.16)}\end{aligned}$$

Similarly, we have for B,

$$\begin{aligned} B = & -\int_0^t \frac{(p(x)dS_0(v \mid x))''\, S(v^- \mid x) - (p(x)S_0(v^- \mid x))''\, dS(v \mid x)}{S^2(v^- \mid x)} \\ = & -\int_0^t \frac{d}{ds}\left[\frac{(p(x)S_0(s \mid x))''}{S(s \mid x)}\right]\Big|_{s=v^-} dv = -\frac{(p(x)S_0(t^- \mid x))''}{S(t \mid x)} + p''(x). \end{aligned} \tag{D.17}$$

Inserting (D.16) and (D.17) in (D.15), we obtain

$$II = \frac{p''(x)}{S(t^- \mid x)} - \frac{(p(x)S_0(t^- \mid x))''}{S(t^- \mid x)} = -\frac{S''(t^- \mid x)}{S(t^- \mid x)}. \tag{D.18}$$

Combining (D.14) and (D.18), the expression for $\Phi_c''(x,t,x)$ is obtained.
As for the second derivative of $\Phi\,(y,t,x)$ in (2.31) evaluated at $y = x$, we have that

$$\begin{aligned} &\Phi''\,(x,t,x) \\ = & -2\int_0^t \frac{G'\,(v^- \mid x)}{1 - G\,(v^- \mid x)} \frac{d}{dy}\left(\frac{dF\,(v \mid y)}{1 - F\,(v^- \mid x)} - \frac{1 - F\,(v^- \mid y)}{(1 - F\,(v^- \mid x))^2} dF\,(v \mid x)\right)\Big|_{y=x} \\ & + \int_0^t \frac{d^2}{dy^2}\left[\frac{dF\,(v \mid y)}{1 - F\,(v^- \mid x)} - \frac{1 - F\,(v^- \mid y)}{(1 - F\,(v^- \mid x))^2} dF\,(v \mid x)\right]\Big|_{y=x} \\ = & \, 2\int_0^t \frac{G'\,(v^- \mid x)}{1 - G\,(v^- \mid x)} \frac{d}{ds}\left(\frac{S'\,(s \mid x)}{S\,(s \mid x)}\right)\Big|_{s=v^-} + II. \end{aligned}$$

Using (D.18), the proof for $\Phi''\,(x,t,x)$ is concluded.

□

Lemma D.5 Let us define $\Phi_1^c(y,t,x) = \mathrm{E}\left(\zeta^2(T,\delta,\xi,\nu,t,x) \mid X=y\right)$, with $\zeta(T,\delta,\xi,\nu,t,x)$ given in (2.21), then

$$\Phi_1^c(x,t,x) = \int_0^t \frac{dH^1(v \mid x)}{J^2(v^- \mid x)}.$$

Proof. By recalling the definition of $\zeta(T,\delta,\xi,\nu,t,x)$ in (2.21), we have

$$\begin{aligned}
&\Phi_1^c(y,t,x)\\
&= \mathrm{E}\left(\frac{\mathbf{1}(T \le t, \delta = 1)}{J^2(T^- \mid x)} \mid X=y\right)\\
&\quad - 2\mathrm{E}\left(\frac{\mathbf{1}(T \le t, \delta = 1)}{J(T^- \mid x)} \int_0^t \frac{\mathbf{1}(T \ge v) + \mathbf{1}(T < v, \xi\nu = 1)}{J^2(v^- \mid x)} dH^1(v \mid x) \mid X=y\right)\\
&\quad + \mathrm{E}\Bigg[\int_0^t \int_0^t \left(\mathbf{1}(T \ge u) + \mathbf{1}(T < u, \xi\nu = 1)\right) \times \left(\mathbf{1}(T \ge v) + \mathbf{1}(T < v, \xi\nu = 1)\right)\\
&\quad \times \frac{dH^1(u \mid x)\, dH^1(v \mid x)}{J^2(u^- \mid x)\, J^2(v^- \mid x)} \mid X=y\Bigg]\\
&= A - 2B + C.
\end{aligned} \tag{D.19}$$

For the first term of (D.19) we have

$$\begin{aligned}
A &= \mathrm{E}\left[\frac{\mathbf{1}(T \le t)}{J^2(T^- \mid x)} \mathrm{E}\left(\mathbf{1}(\delta = 1) \mid T, X=y\right) \mid X=y\right]\\
&= \mathrm{E}\left(\frac{\mathbf{1}(T \le t)}{J^2(T^- \mid x)} \phi(T,y) \mid X=y\right)\\
&= \int_0^t \frac{\phi(v,y)}{J^2(v^- \mid x)} dH(v \mid y) = \int_0^t \frac{dH^1(v \mid y)}{J^2(v^- \mid x)},
\end{aligned}$$

where $\phi(v,y) = E(\delta \mid T=v, X=y)$.

For the second term in (D.19),

$$\begin{aligned}
B &= \mathrm{E}\Bigg[\int_0^t \frac{1}{J^2(v^- \mid x)} \frac{\mathbf{1}(T \le t, \delta = 1)\left(\mathbf{1}(T \ge v) + \mathbf{1}(T < v, \xi\nu = 1)\right)}{J(T^- \mid x)}\\
&\quad \times dH^1(v \mid x) \mid X=y\Bigg]\\
&= \int_0^t \frac{1}{J^2(v^- \mid x)} \mathrm{E}\left[\frac{\mathbf{1}(T \le t, \delta = 1)\left(\mathbf{1}(T \ge v) + \mathbf{1}(T < v, \xi\nu = 1)\right)}{J(T^- \mid x)} \mid X=y\right]\\
&\quad \times dH^1(v \mid x)\\
&= \int_0^t \frac{1}{J^2(v^- \mid x)} \mathrm{E}\left(\frac{\mathbf{1}(v \le T \le t, \delta = 1)}{J(T^- \mid x)} \mid X=y\right) dH^1(v \mid x)\\
&= \int_0^t \frac{1}{J^2(v^- \mid x)} \left(\int_v^t \frac{dH^1(u \mid y)}{J(u^- \mid x)}\right) dH^1(v \mid x).
\end{aligned}$$

Finally, for the third term in (D.19),

$$\begin{aligned}C &= \mathrm{E}\left(\int_0^t\int_0^t \frac{\mathbf{1}\left(T\geq u\right)\mathbf{1}\left(T\geq v\right)}{J^2\left(u^-\mid x\right)J^2\left(v^-\mid x\right)}dH^1\left(u\mid x\right)dH^1\left(v\mid x\right)\mid X=y\right)\\ &+\mathrm{E}\left(\int_0^t\int_0^t \frac{\mathbf{1}\left(T\geq u\right)\mathbf{1}\left(T< v,\xi\nu=1\right)}{J^2\left(u^-\mid x\right)J^2\left(v^-\mid x\right)}dH^1\left(u\mid x\right)dH^1\left(v\mid x\right)\mid X=y\right)\\ &+\mathrm{E}\left(\int_0^t\int_0^t \frac{\mathbf{1}\left(T< u,\xi\nu=1\right)\mathbf{1}\left(T\geq v\right)}{J^2\left(u^-\mid x\right)J^2\left(v^-\mid x\right)}dH^1\left(u\mid x\right)dH^1\left(v\mid x\right)\mid X=y\right)\\ &+\mathrm{E}\left(\int_0^t\int_0^t \frac{\mathbf{1}\left(T< u,\xi\nu=1\right)\mathbf{1}\left(T< v,\xi\nu=1\right)}{J^2\left(u^-\mid x\right)J^2\left(v^-\mid x\right)}dH^1\left(u\mid x\right)dH^1\left(v\mid x\right)\mid X=y\right)\\ &= C_1+C_2+C_3+C_4. \qquad (D.20)\end{aligned}$$

The first expectation in (D.20) is

$$\begin{aligned}C_1 &= \int_0^t\int_0^t \frac{\mathrm{E}\left(\mathbf{1}\left(T\geq\max\left(u,v\right)\right)\mid X=y\right)}{J^2\left(u^-\mid x\right)J^2\left(v^-\mid x\right)}dH^1\left(u\mid x\right)dH^1\left(v\mid x\right)\\ &= \int_0^t\int_0^t \frac{1-H\left(\max\left(u,v\right)^-\mid y\right)}{J^2\left(u^-\mid x\right)J^2\left(v^-\mid x\right)}dH^1\left(u\mid x\right)dH^1\left(v\mid x\right).\end{aligned}$$

Integrating on the supports $(u,v)\in[0,t]\times[0,t]\,,u\leq v$, and $(u,v)\in[0,t]\times[0,t]\,,u>v$, we have

$$C_1 = 2\int_0^t \frac{1}{J^2\left(u^-\mid x\right)}\left(\int_u^t \frac{1-H\left(v^-\mid y\right)}{J^2\left(v^-\mid x\right)}dH^1\left(v\mid x\right)\right)dH^1\left(u\mid x\right). \qquad (D.21)$$

For the second and third terms in (D.20), we get

$$\begin{aligned}C_2=C_3 &= \int_0^t \frac{1}{J^2\left(u^-\mid x\right)}\int_0^t \frac{\mathrm{E}\left(\mathbf{1}\left(u\leq T<v,\xi\nu=1\right)\mid X=y\right)}{J^2\left(v^-\mid x\right)}dH^1\left(v\mid x\right)dH^1\left(u\mid x\right)\\ &= \int_0^t \frac{1}{J^2\left(u^-\mid x\right)}\int_u^t \frac{H^{11}\left(v^-\mid y\right)-H^{11}\left(u^-\mid y\right)}{J^2\left(v^-\mid x\right)}dH^1\left(v\mid x\right)dH^1\left(u\mid x\right)\\ &= \int_0^t \frac{1}{J^2\left(u^-\mid x\right)}\left(\int_u^t \frac{H^{11}\left(v^-\mid y\right)}{J^2\left(v^-\mid x\right)}dH^1\left(v\mid x\right)\right)dH^1\left(u\mid x\right)\\ &- \int_0^t \frac{H^{11}\left(u^-\mid y\right)}{J^2\left(u^-\mid x\right)}\left(\int_u^t \frac{dH^1\left(v\mid x\right)}{J^2\left(v^-\mid x\right)}\right)dH^1\left(u\mid x\right). \qquad (D.22)\end{aligned}$$

Finally, for the last term in (D.20), we have

$$\begin{aligned}C_4 &= \int_0^t\int_0^t \frac{\mathrm{E}\left(\mathbf{1}\left(T<\min\left(u,v\right),\xi\nu=1\right)\mid X=y\right)}{J^2\left(u^-\mid x\right)J^2\left(v^-\mid x\right)}dH^1\left(u\mid x\right)dH^1\left(v\mid x\right)\\ &= \int_0^t\int_0^t \frac{H^{11}\left(\min\left(u,v\right)^-\mid y\right)}{J^2\left(u^-\mid x\right)J^2\left(v^-\mid x\right)}dH^1\left(u\mid x\right)dH^1\left(v\mid x\right).\end{aligned}$$

Integrating on the supports $(u,v)\in[0,t]\times[0,t]\,,u\geq v$ and $(u,v)\in[0,t]\times[0,t]\,,v>u$, we have

$$\begin{aligned}C_4 = &\int_0^t \frac{H^{11}(v^- \mid y)}{J^2(v^- \mid x)} \left(\int_v^t \frac{dH^1(u \mid x)}{J^2(u^- \mid x)} \right) dH^1(v \mid x) \\ &+ \int_0^t \frac{H^{11}(u^- \mid y)}{J^2(u^- \mid x)} \left(\int_u^t \frac{dH^1(v \mid x)}{J^2(v^- \mid x)} \right) dH^1(u \mid x) \\ &= 2\int_0^t \frac{H^{11}(v^- \mid y)}{J^2(v^- \mid x)} \left(\int_v^t \frac{dH^1(u \mid x)}{J^2(u^- \mid x)} \right) dH^1(v \mid x). \qquad (D.23)\end{aligned}$$

Combining (D.21), (D.22) and (D.23), we obtain

$$\begin{aligned}C &= 2\int_0^t \frac{1}{J^2(u^- \mid x)} \left(\int_u^t \frac{1 - H(v^- \mid y) + H^{11}(v^- \mid y)}{J^2(v^- \mid x)} dH^1(v \mid x) \right) dH^1(u \mid x) \\ &= 2\int_0^t \frac{1}{J^2(u^- \mid x)} \left(\int_u^t \frac{J(v^- \mid y)}{J^2(v^- \mid x)} dH^1(v \mid x) \right) dH^1(u \mid x).\end{aligned}$$

Note that for $y = x$, then $C = 2B$. Therefore, after doing the sum in (D.19), we get

$$\Phi_1^c(x,t,x) = \int_0^t \frac{dH^1(v \mid x)}{J^2(v^- \mid x)}.$$

This concludes the proof. □

Lemma D.6 Let X be a continuous random variable with density function $m(x)$ fulfilling Assumptions 2(i), and the kernel K fulfills Assumption 8. Then, for $g(x)$ a four times continuously differentiable function,

$$E(K_h(x - X)g(X)) = m(x)g(x) + \frac{h^2}{2}(g(x)m(x))'' d_K + O\left(h^4\right)$$

where $d_K = \int v^2 K(v)\,dv$.

Proof. A change of variable and a Taylor expansion with the remainder term in integral form yield

$$\begin{aligned}&E(K_h(x - X)g(X)) \\ &= \int \frac{1}{h}K\left(\frac{x-u}{h}\right) g(u)m(u)\,du = \int K(v)g(x - vh)m(x - vh)\,dv \\ &= \int K(v) \left[g(x)m(x) - vh(g(x)m(x))' + \frac{v^2h^2}{2}(g(x)m(x))'' - \frac{v^3h^3}{6}(g(x)m(x))''' \right. \\ &\quad \left. + \frac{1}{6}\int_x^{x-vh} (x - vh - w)^3 (g(w)m(w))^{(4)} dw \right] dv \\ &= m(x)g(x) + \frac{h^2}{2}(g(x)m(x))'' d_K\end{aligned}$$

$$+\frac{1}{6}\int K\left(v\right)\int_{x}^{x-vh}\left(x-vh-w\right)^{3}\left(g\left(w\right)m\left(w\right)\right)^{(4)}dwdv.$$

With a new change of variable in the remainder term and a Taylor expansion it can be seen that

$$\begin{aligned}
&E\left(K_{h}\left(x-X\right)g\left(X\right)\right)\\
&=m\left(x\right)g\left(x\right)+\frac{h^{2}}{2}\left(g\left(x\right)m\left(x\right)\right)''d_{K}\\
&\quad+\frac{1}{6}\int K\left(v\right)\int_{0}^{1}\left(-vh\right)^{3}\left(1-t\right)^{3}\left(g\left(x-vht\right)\left(-vh\right)m\left(x-vht\right)\left(-vh\right)\right)^{(4)}dtdv\\
&=m\left(x\right)g\left(x\right)+\frac{h^{2}}{2}\left(g\left(x\right)m\left(x\right)\right)''d_{K}\\
&\quad-\frac{1}{6}\int v^{4}h^{4}K\left(v\right)\int_{0}^{1}\left(1-t\right)^{3}\left(g\left(x-vht\right)m\left(x-vht\right)\right)^{(4)}dtdv\\
&=m\left(x\right)g\left(x\right)+\frac{h^{2}}{2}\left(g\left(x\right)m\left(x\right)\right)''d_{K}+O\left(h^{4}\right).
\end{aligned}$$

This completes the proof. □

Lemma D.7 Let X be a continuous random variable with density function $m(x)$ fulfilling Assumptions 2(i), and the kernel K fulfills Assumption 8. Then, for $g\left(x\right)$ a four times continuously differentiable function,

$$E\left(K_{h}^{2}\left(x-X\right)g\left(X\right)\right)=\frac{1}{h}m\left(x\right)g\left(x\right)c_{K}+\frac{1}{2}\left(g\left(x\right)m\left(x\right)\right)''e_{K}h+O\left(h^{3}\right)$$

where $c_{K}=\int K^{2}\left(v\right)dv$ and $e_{K}=\int v^{2}K^{2}\left(v\right)dv$.

Proof. A change of variable and a Taylor expansion yield

$$\begin{aligned}
&E\left(K_{h}^{2}\left(x-X\right)g\left(X\right)\right)\\
&=\int\frac{1}{h^{2}}K^{2}\left(\frac{x-u}{h}\right)g\left(u\right)m\left(u\right)du=\frac{1}{h}\int K^{2}\left(v\right)g\left(x-vh\right)m\left(x-vh\right)dv\\
&=\frac{1}{h}\int K^{2}\left(v\right)\Bigg(g\left(x\right)m\left(x\right)-vh\left(g\left(x\right)m\left(x\right)\right)'+\frac{v^{2}h^{2}}{2}\left(g\left(x\right)m\left(x\right)\right)''\\
&\quad-\frac{v^{3}h^{3}}{6}\left(g\left(x\right)m\left(x\right)\right)'''-\frac{1}{6}\frac{1}{h}\int v^{4}h^{4}K^{2}\left(v\right)\\
&\quad\times\int_{0}^{1}\left(1-t\right)^{3}\left(g\left(x-vht\right)m\left(x-vht\right)\right)^{(4)}dt\Bigg)dv\\
&=\frac{1}{h}m\left(x\right)g\left(x\right)c_{K}+\frac{1}{2}\left(g\left(x\right)m\left(x\right)\right)''e_{K}h+O\left(h^{3}\right).
\end{aligned}$$

This completes the proof. □

Lemma D.8 Let X be a continuous random variable with density function $m(x)$ fulfilling Assumptions 2(i), and the kernel K fulfils Assumption 8. Let $g_1(x)$ and $g_2(x)$ be four times continuously differentiable functions, then

$$\begin{aligned}
& E\left(K_h(x-X_1)\,K_h(x-X_2)\,g_1(X_1)\,g_2(X_2)\right) \\
={}& g_1(x)\,g_2(x)\,m^2(x) \\
& +\frac{1}{2}h^2 m(x)\left(g_1(x)\left(g_2(x)\,m(x)\right)''+\left(g_1(x)\,m(x)\right)''g_2(x)\right)d_K+O\left(h^4\right)
\end{aligned}$$

where $d_K=\int v^2K(v)\,dv$.

Proof. A change of variable and a Taylor expansion yield

$$\begin{aligned}
& E\left(K_h(x-X_1)\,K_h(x-X_2)\,g_1(X_1)\,g_2(X_2)\right) \\
={}& \iint \frac{1}{h^2}K\left(\frac{x-u_1}{h}\right)K\left(\frac{x-u_2}{h}\right)g_1(u_1)\,g_2(u_2)\,m(u_1)\,m(u_2)\,du_1du_2 \\
={}& \iint K(v_1)\,K(v_2)\,g_1(x-v_1h)\,g_2(x-v_2h)\,m(x-v_1h)\,m(x-v_2h)\,dv_1dv_2 \\
={}& \iint K(v_1)\,K(v_2)\Bigg[g_1(x)\,m(x)-v_1h\left(g_1(x)\,m(x)\right)'+\frac{v_1^2h^2}{2}\left(g_1(x)\,m(x)\right)'' \\
& -\frac{v_1^3h^3}{6}\left(g_1(x)\,m(x)\right)'''-\frac{v_1^4h^4}{6}\int_0^1(1-t_1)^3\left(g_1(x-v_1ht_1)\,m(x-v_1ht_1)\right)^{(4)}dt_1\Bigg] \\
& \times\Bigg[g_2(x)\,m(x)-v_2h\left(g_2(x)\,m(x)\right)'+\frac{1}{v_2^2h^2}\left(g_2(x)\,m(x)\right)'' \\
& -\frac{1}{v_2^3h^3}\left(g_2(x)\,m(x)\right)'''-\frac{v_2^4h^4}{6}\int_0^1(1-t_2)^3\left(g_2(x-v_2ht_2)\,m(x-v_2ht_2)\right)^{(4)}dt_2\Bigg] \\
={}& g_1(x)\,g_2(x)\,m^2(x)+\frac{1}{2}h^2m(x)\left[g_1(x)\left(g_2(x)\,m(x)\right)''d_K\right. \\
& \left.+\left(g_1(x)\,m(x)\right)''g_2(x)\right]d_K+O\left(h^4\right).
\end{aligned}$$

This completes the proof. □

Lemma D.9 Consider the notation in the proof of Proposition 3.5. It can be proved that the expectation

$$E\left[(1-\xi_1)\,E\left(\widehat{\mu}_1^*(X_1,g_2)\mid O\right)^2\right]=E\left[(1-\pi(X))\,\mu^2(X)\right]\left(1+o\left(g_1^2+g_2^2\right)\right).$$

Proof. Start by defining the following expectations conditioned on the observed data and the bootstrap resamples:

$$E\left(\nu_1^+\mid O,R\right)=\widehat{\mu}_1^*(X_1,g_2)=\sum_{i=1}^n B_{g_2i}(X_1)\,\nu_i^*$$

$$E\left(\nu_1^{+2} \mid O, R\right) = \widehat{\mu}_2^*(X_1, g_2) = \sum_{i=1}^{n} B_{g_2 i}(X_1)\, \nu_i^{*2}.$$

Then,

$$\begin{aligned} E\left(\widehat{\mu}_1^*(X_1, g_2) \mid O\right) &= E\left(\sum_{i=1}^{n} B_{g_2 i}(X_1)\, \nu_i^* \mid O\right) = \sum_{i=1}^{n} B_{g_2 i}(X_1) \sum_{j=1}^{n} B_{g_1 j}(X_i)\, \nu_j \\ E\left(\widehat{\mu}_2^*(X_1, g_2) \mid O\right) &= E\left(\sum_{i=1}^{n} B_{g_2 i}(X_1)\, \nu_i^{*2} \mid O\right) = \sum_{i=1}^{n} B_{g_2 i}(x) \sum_{j=1}^{n} B_{g_1 j}(X_i)\, \nu_j^2. \end{aligned}$$

From (B.18), it can be shown that

$$\begin{aligned} & E\left(\widehat{\mu}_1^*(x, g_2) \mid O\right) \\ \simeq & \frac{1}{n^2 g_1 g_2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\pi^{-1}(X_i)\, m^{-1}(X_i)}{\pi(X_1)\, m(X_1)} \xi_i \xi_j K\left(\frac{X_1 - X_i}{g_2}\right) K\left(\frac{X_i - X_j}{g_1}\right) \nu_j, \end{aligned}$$

and similarly

$$\begin{aligned} & E\left(\sum_{i=1}^{n} B_i(X_1, g_2)\, \nu_i^{*2} \mid O\right) \\ \simeq\ & \frac{1}{n^2 g_1 g_2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\pi^{-1}(X_i)\, m^{-1}(X_i)}{\pi(X_1)\, m(X_1)} \xi_i \xi_j K\left(\frac{X_1 - X_i}{g_2}\right) K\left(\frac{X_i - X_j}{g_1}\right) \nu_j^2. \end{aligned}$$

Note that

$$\begin{aligned} & E\left[(1-\xi_1)\, E\left(\widehat{\mu}_1^*(X_1, g_2) \mid O\right)^2\right] \\ = & E\left[(1-\xi_1)\, E\left(\sum_{i=1}^{n} B_{g_1 i}(X_1)\, \nu_i^* \mid O\right)^2\right] \\ \simeq & \frac{1}{n^4 g_1^2 g_2^2} E\left[(1-\xi_1) \frac{\pi^{-2}(X_1)}{m^2(X_1)} \left(\sum_{i=2}^{n} \sum_{j=2}^{n} \frac{\pi^{-1}(X_i)}{m(X_i)} \xi_i \xi_j K\left(\frac{X_1 - X_i}{g_2}\right) K\left(\frac{X_j - X_i}{g_1}\right) \nu_j\right)^2\right] \\ = & \frac{1}{n^4 g_1^2 g_2^2} E\left[(1-\xi_1) \frac{\pi^{-2}(X_1)}{m^2(X_1)} \sum_{i=2}^{n} \sum_{j=2}^{n} \sum_{k=2}^{n} \sum_{l=2}^{n} \frac{\pi^{-1}(X_i)}{m(X_i)} \frac{\pi^{-1}(X_k)}{m(X_k)} \xi_i \xi_j \xi_k \xi_l \right. \\ & \left. K\left(\frac{X_1 - X_i}{g_2}\right) K\left(\frac{X_j - X_i}{g_1}\right) K\left(\frac{X_1 - X_k}{g_2}\right) K\left(\frac{X_l - X_k}{g_1}\right) \nu_j \nu_l\right]. \end{aligned}$$

There are 15 cases to be considered:

(a1) $i = j = k = l$

(a2) $i = j = k$ and $l \neq i$

(a3) $i = j, k \neq i$ and $l = i$

(a4) $i = j, k \neq i$ and $l = j$

(a5) $i = j, k \neq i$ and $l \neq i, l \neq j$

(a6) $j \neq i, k = l = i$

(a7) $j \neq i, k = i$ and $l = j$

(a8) $j \neq i, k = i$ and $l \neq i$ and $l \neq j$

(a9) $j \neq i, k = j$ and $l = i$

(a10) $j \neq i, k = j$ and $l = j$

(a11) $j \neq i, k = j$ and $l \neq i$ and $l \neq j$

(a12) $j \neq i, k \neq i$ and $k \neq j$, and $l = i$

(a13) $j \neq i, k \neq i$ and $k \neq j$, and $l = j$

(a14) $j \neq i, k \neq i$ and $k \neq j$, and $l = k$

(a15) $j \neq i, k \neq i$ and $k \neq j$, and $l \neq i$,$l \neq j$ and $l \neq k$.

Case (a1): $i = j = k = l$

$$
\begin{aligned}
& E\left((1-\xi_1) \frac{\pi^{-2}(X_1)}{m^2(X_1)} \frac{\pi^{-1}(X_2)}{m(X_2)} \frac{\pi^{-1}(X_2)}{m(X_2)} \xi_2 K^2\left(\frac{X_1 - X_2}{g_2}\right) K^2(0)\, \nu_2^2 \right) \\
&= K^2(0) \iint \frac{\pi^{-2}(x_1)}{m^2(x_1)} \frac{\pi^{-1}(x_2)}{m(x_2)} \frac{\pi^{-1}(x_2)}{m(x_2)} (1-\pi(x_1))\, \pi(x_2)\, K^2\left(\frac{x_1 - x_2}{g_2}\right) \\
&\quad \times \left(\sigma^2(x_2) + \mu^2(x_2)\right) m(x_1)\, m(x_2)\, dx_1 dx_2 \\
&= K^2(0) \iint \frac{\pi^{-2}(x_1)}{m(x_1)} \frac{\pi^{-1}(x_2)}{m(x_2)} (1-\pi(x_1)) \\
&\quad \times K^2\left(\frac{x_1 - x_2}{g_2}\right) \left(\sigma^2(x_2) + \mu^2(x_2)\right) dx_1 dx_2.
\end{aligned}
$$

Applying a change of variable and a Taylor expansion,

$$
\begin{aligned}
& g_2 K^2(0) \iint \frac{\pi^{-2}(x_2 + g_1 v_1)}{m(x_2 + g_1 v_1)} \frac{\pi^{-1}(x_2)}{m(x_2)} (1 - \pi(x_2 + g_1 v_1)) \\
&\quad \times K^2(v_1) \left(\sigma^2(x_2) + \mu^2(x_2)\right) dv_1 dx_2 \\
&= g_2 K^2(0)\, c_K \int \frac{\pi^{-2}(x_2)}{m(x_2)} \frac{\pi^{-1}(x_2)}{m(x_2)} (1 - \pi(x_2)) \left(\sigma^2(x_2) + \mu^2(x_2)\right) dx_2 \left(1 + g_2^2\right) \\
&= O(g_2),
\end{aligned}
$$

where $c_K = \int K^2(v)\, dv$. Applying similar ideas, the only dominant term is given by case (a15) as Cases (a1) – (a14) are negligible. For example,

Case (a7): $i \neq j, k = i, l \neq i, l = j$

$$
(n-1)(n-2) \iiint (1 - \pi(x_1)) \frac{\pi^{-2}(X_1)}{m(X_1)} \frac{\pi^{-1}(X_2)}{} \frac{}{m(X_2)} \pi(x_3)
$$

$$\begin{aligned}
& \times K^2\left(\frac{X_1-X_2}{g_2}\right)K^2\left(\frac{X_3-X_2}{g_1}\right)\mu^2\left(x_3\right)m\left(x_3\right)dx_1dx_2dx_3 \\
=\ & (n-1)(n-2)\,g_1g_2c_K^2\int\left(1-\pi\left(x_2\right)\right)\frac{\pi^{-2}\left(x_2\right)}{m\left(x_2^2\right)}\mu^2\left(x_2\right)m\left(x_2\right)dx_2\left(1+o(1)\right) \\
=\ & n^2g_1g_2c_K^2E\left(\left(1-\pi(X)\right)\frac{\pi^{-2}(X)}{m(X)}\mu^2(X)\right)\left(1+o(1)\right) \\
=\ & O\left(n^2g_1g_2\right).
\end{aligned}$$

Case (a8): $i\neq j, k=i, l\neq i, l\neq j$

$$\begin{aligned}
& (n-1)(n-2)(n-3) \\
& \times\iiiint\Bigg[\left(1-\pi\left(x_1\right)\right)\frac{\pi^{-2}\left(X_1\right)}{m\left(X_1\right)}\frac{\pi^{-1}\left(X_2\right)}{m\left(X_2\right)}\frac{\pi^{-1}\left(X_2\right)}{}\pi\left(x_2\right)\pi\left(x_3\right)\pi\left(x_4\right) \\
& \times K^2\left(\frac{X_1-X_2}{g_2}\right)K\left(\frac{X_3-X_2}{g_1}\right)K\left(\frac{X_4-X_2}{g_1}\right)\mu\left(x_3\right)\mu\left(x_4\right)m\left(x_3\right)m\left(x_4\right) \\
& \times dx_1dx_2dx_3dx_4] \\
=\ & (n-1)(n-2)(n-3)\,g_1^2g_2c_K \\
& \times\int\left(1-\pi\left(x_2\right)\right)\pi^{-1}\left(X_2\right)\frac{\mu^2\left(x_2\right)}{m\left(X_2\right)}m\left(x_2\right)dx_2\left(1+o(1)\right) \\
=\ & n^3g_1^2g_2c_KE\left(\left(1-\pi(X)\right)\pi^{-1}(X)\frac{\mu^2(X)}{m(X)}\right)\left(1+o(1)\right).
\end{aligned}$$

Case (a9): $i\neq j, k=j, l=i$

$$\begin{aligned}
& (n-1)(n-2)\iiint\left(1-\pi\left(x_1\right)\right)\frac{\pi^{-2}\left(X_1\right)}{m\left(X_1\right)} \\
& \times K\left(\frac{X_1-X_2}{g_2}\right)K^2\left(\frac{X_3-X_2}{g_1}\right)K\left(\frac{X_1-X_3}{g_2}\right)\mu\left(x_3\right)\mu\left(x_2\right)dx_1dx_2dx_3 \\
=\ & (n-1)(n-2)\,g_2^2\iiint\left(1-\pi\left(x_1\right)\right)\frac{\pi^{-2}\left(X_1\right)}{m\left(X_1\right)} \\
& \times K\left(v_1\right)K^2\left(\frac{g_2}{g_1}\left(v_3-v_1\right)\right)K\left(v_3\right)\mu^2\left(x_1\right)dx_1dv_2dv_3\left(1-o(1)\right) \\
=\ & O\left(n^2g_2^2\right).
\end{aligned}$$

Case (a11): $i\neq j,\ k=j, l\neq j, l\neq i$

$$\begin{aligned}
& (n-1)(n-2)(n-3)\iiint\left(1-\pi\left(x_1\right)\right)\frac{\pi^{-2}\left(X_1\right)}{m\left(X_1\right)} \\
& \times\pi\left(x_4\right)K\left(\frac{X_1-X_2}{g_2}\right)K\left(\frac{X_3-X_2}{g_1}\right)K\left(\frac{X_1-X_3}{g_2}\right)K\left(\frac{X_4-X_3}{g_2}\right) \\
& \times\mu\left(x_3\right)\mu\left(x_4\right)m\left(x_4\right)dx_1dx_2dx_3dx_4
\end{aligned}$$

$$
\begin{aligned}
&= (n-1)(n-2)(n-3)\, g_2^2 g_1 \iiint (1-\pi(x_1)) \frac{\pi^{-1}(X_1)}{m(X_1)} \\
&\quad \times K(v_1) K(v_3) K\left(v_{1-}\frac{g_1}{g_2}v_3\right) K(v_4)\, \mu(x_3)\, \mu(x_4)\, m(x_4)\, dx_1 dx_2 dx_3 dx_4 \\
&= O\left(n^3 g_2^2 g_1\right).
\end{aligned}
$$

Case (a15): $i \neq j, k \neq i, k \neq j, l \neq k, l \neq j, l \neq i$

$$
\begin{aligned}
& (n-1)(n-2)(n-3)(n-4) \int \cdots \int (1-\pi(x_1)) \frac{\pi^{-2}(x_1)}{m^2(x_1)} \frac{\pi^{-1}(x_2)}{m(x_2)} \frac{\pi^{-1}(x_4)}{m(x_4)} \pi(x_2) \\
& \times \pi(x_3)\, \pi(x_4)\, \pi(x_5)\, K\left(\frac{X_1 - X_2}{g_2}\right) K\left(\frac{X_3 - X_2}{g_1}\right) K\left(\frac{X_1 - X_4}{g_2}\right) K\left(\frac{X_5 - X_4}{g_1}\right) \\
& \times \mu(x_3)\, \mu(x_5)\, m(x_1)\, m(x_2)\, m(x_3)\, m(x_4)\, m(x_5)\, dx_1 dx_2 dx_3 dx_4 dx_5 \\
= \; & (n-1)(n-2)(n-3)(n-4) \int \cdots \int (1-\pi(x_1)) \frac{\pi^{-2}(x_1)}{m(x_1)} \pi(x_3)\, \pi(x_5) \\
& \times K\left(\frac{X_1 - X_2}{g_2}\right) K\left(\frac{X_3 - X_2}{g_1}\right) K\left(\frac{X_1 - X_4}{g_2}\right) K\left(\frac{X_5 - X_4}{g_1}\right) \\
& \times \mu(x_3)\, \mu(x_5)\, m(x_3)\, m(x_5)\, dx_1 dx_2 dx_3 dx_4 dx_5 \\
= \; & (n-1)(n-2)(n-3)(n-4)\, g_1^2 g_2^2 \int \cdots \int (1-\pi(x_1)) \frac{\pi^{-2}(x_1)}{m(x_1)} \\
& \times K(v_2) K(v_3) K(v_4) K(v_5)\, \pi(x_1 - g_2 v_2 + g_1 v_3)\, \pi(x_1 - g_2 v_4 + g_1 v_5) \\
& \times \mu(x_1 - g_2 v_2 + g_1 v_3)\, \mu(x_1 - g_2 v_4 + g_1 v_5)\, m(x_1 - g_2 v_2 + g_1 v_3) \\
& \times m(x_1 - g_2 v_4 + g_1 v_5)\, dx_1 dv_2 dv_3 dv_4 dv_5 \\
= \; & (n-1)(n-2)(n-3)(n-4)\, g_1^2 g_2^2 \int \cdots \int (1-\pi(x_1)) \frac{\pi^{-2}(x_1)}{m(x_1)} \\
& \times \left(\pi^2(x_1)\, \mu^2(x_1)\, m^2(x_1) + O\left(g_1^2 + g_2^2\right)\right) K(v_2) K(v_3) K(v_4) \\
& \times K(v_5)\, dx_1 dv_2 dv_3 dv_4 dv_5 \\
= \; & (n-1)(n-2)(n-3)(n-4)\, g_1^2 g_2^2 \left(E\left[(1-\pi(X))\, \mu^2(X)\right] + O\left(g_1^2 + g_2^2\right)\right) \\
= \; & n^4 g_1^2 g_2^2 E\left[(1-\pi(X))\, \mu^2(X)\right] + O\left(n^4 g_1^2 g_2^2 \left(g_1^2 + g_2^2\right)\right).
\end{aligned}
$$

Rejoining the cases, the proof is completed.

□

Lemma D.10 Consider the notation in the proof of Proposition 3.5. It can be shown that the expectation

$$
\begin{aligned}
E\left[(1-\xi_1)\, E\left(\nu_1^{+2} \mid O, R\right)\right] &= E\left[(1-\xi_1)\, \widehat{\mu}_2^*(X_1, g_2)\right] \\
&= E\left[(1-\pi(X))\left(\sigma^2(X) + \mu^2(X)\right)\right]\left(1 + o\left(g_1^2 + g_2^2\right)\right).
\end{aligned}
$$

Proof. Note that

$$\begin{aligned} E\left[(1-\xi_1)\,E\left(\nu_1^{+2}\mid O,R\right)\right] &= E\left[(1-\xi_1)\,\widehat{\mu}_2^{*}\left(X_1,g_2\right)\right] \\ &= E\left[(1-\xi_1)\,E\left(\sum_{i=1}^{n} B_{g_2 i}\left(X_1\right)\nu_i^{*2}\mid O\right)\right]. \end{aligned}$$

The inner expectation is

$$\begin{aligned} & E\left(\sum_{i=1}^{n} w_{g_2 i}\left(X_1\right)\nu_i^{*2}\mid O\right) \\ & \simeq \frac{1}{n^2 g_1 g_2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\pi^{-1}\left(X_i\right)m^{-1}\left(X_i\right)}{\pi\left(X_1\right)m\left(X_1\right)}\xi_i\xi_j K\left(\frac{X_1-X_i}{g_2}\right)K\left(\frac{X_i-X_j}{g_1}\right)\nu_j^2. \end{aligned}$$

Different cases must be considered according to the values of i and j. Note that $(1-\xi_1)\,\xi_i\xi_j = 0$ if $i=1$ or $j=i$. So only two cases to be considered:

Case (a): $i \neq 1$, $j \neq 1$ but $i = j$

$$\begin{aligned} & E\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\left(\pi\left(X_1\right)\pi\left(X_i\right)\right)^{-1}}{m\left(X_1\right)m\left(X_i\right)}\left(1-\xi_1\right)\xi_i\xi_j K\left(\frac{X_1-X_i}{g_2}\right)K\left(\frac{X_j-X_i}{g_1}\right)\nu_j^2\right) \\ = \; & (n-1)\,K(0)\iint\frac{\left(\pi\left(X_1\right)\pi\left(X_2\right)\right)^{-1}}{m\left(X_1\right)m\left(X_2\right)}\left(1-\pi\left(x_1\right)\right)\pi\left(x_2\right)K\left(\frac{X_1-X_2}{g_2}\right) \\ & \times\mu^2\left(x_2\right)m\left(x_1\right)m\left(x_2\right)dx_1dx_2 \\ = \; & (n-1)\,K(0)\iint\frac{1-\pi\left(x_1\right)}{\pi\left(x_1\right)}K\left(\frac{X_1-X_2}{g_2}\right)\mu^2\left(x_2\right)dx_1dx_2 = O\left(ng_2\right). \end{aligned}$$

Case (b): $i \neq 1$, $j \neq 1$ but $i \neq j$

$$\begin{aligned} & E\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\left(\pi\left(X_1\right)\pi\left(X_i\right)\right)^{-1}}{m\left(X_1\right)m\left(X_i\right)}\left(1-\xi_1\right)\xi_i\xi_j K\left(\frac{X_1-X_i}{g_2}\right)K\left(\frac{X_j-X_i}{g_1}\right)\nu_j^2\right) \\ = \; & (n-1)(n-2)\iint\frac{\left(\pi\left(X_1\right)\pi\left(X_2\right)\right)^{-1}}{m\left(X_1\right)m\left(X_2\right)}\left(1-\pi\left(x_1\right)\right)\pi\left(x_2\right)\pi\left(x_3\right) \\ & K\left(\frac{X_1-X_2}{g_2}\right)K\left(\frac{X_3-X_2}{g_1}\right)m\left(x_1\right)m\left(x_2\right)m\left(x_3\right)\left(\sigma^2\left(x_3\right)+\mu^2\left(x_3\right)\right)dx_1dx_2dx_3 \\ = \; & (n-1)(n-2)\iint\frac{1-\pi\left(x_1\right)}{\pi\left(x_1\right)}\pi\left(x_3\right)K\left(\frac{X_1-X_2}{g_2}\right)K\left(\frac{X_3-X_2}{g_1}\right) \\ & \times m\left(x_3\right)\left(\sigma^2\left(x_3\right)+\mu^2\left(x_3\right)\right)dx_1dx_2dx_3 \\ = \; & (n-1)(n-2)\,g_1g_2E\left(\left(1-\pi(X)\right)\left(\sigma^2(X)+\mu^2(X)\right)\right)\left(1+o\left(g_1^2+g_2^2\right)\right) \\ = \; & n^2g_1g_2E\left(\left(1-\pi(X)\right)\left(\sigma^2(X)+\mu^2(X)\right)\right)\left(1+o\left(g_1^2+g_2^2\right)\right). \end{aligned}$$

Rejoining cases (a) and (b), the proof is completed. □

Lemma D.11 Consider the notation in the proof of Proposition 3.5. It can be shown that the expectation

$$E\left[(1-\xi_1)\,E\left(\widehat{\mu}_1^*\left(X_1,g_2\right)^2\mid O\right)\right]=E\left[(1-\pi(X))\,\mu^2(X)\right]\left(1+o\left(g_1^2+g_2^2\right)\right).$$

Proof. The expectation is

$$\begin{aligned}
&E\left[(1-\xi_1)\,E\left(\widehat{\mu}_1^*\left(X_1,g_2\right)^2\mid O\right)\right]\\
&=E\left[(1-\xi_1)\,E\left(\left(\sum_{i=1}^{n}B_{g_2i}(X_1)\,\nu_i^*\right)^2\mid O\right)\right]\\
&=E\left[(1-\xi_1)\,E\left(\sum_{i=1}^{n}\sum_{j=1}^{n}B_{g_2i}(X_1)\,B_{g_2j}(X_1)\,\nu_i^*\nu_j^*\mid O\right)\right]\\
&=E\left[(1-\xi_1)\sum_{\substack{i=1\\ j=i}}^{n}B_{g_2i}^2(X_1)\,E\left(\nu_i^{*2}\mid O\right)\right]\\
&\quad+E\left[(1-\xi_1)\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}B_{g_2i}(X_1)\,B_{g_2j}(X_1)\,E\left(\nu_i^*\nu_j^*\mid O\right)\right]\\
&=E\left[(1-\xi_1)\sum_{i=1}^{n}B_{g_2i}^2(X_1)\sum_{k=1}^{n}B_{g_1k}(X_i)\,\nu_k^2\right]\\
&\quad+(n-1)(n-2)\,E\left[(1-\xi_1)\,B_{g_22}(X_1)\,B_{g_23}(X_1)\,E\left(\nu_2^*\mid O\right)E\left(\nu_3^*\mid O\right)\right].
\end{aligned}$$

The first expectation is neglibigle,

$$\begin{aligned}
&E\left[(1-\xi_1)\sum_{i=1}^{n}B_{g_2i}^2(X_1)\sum_{k=1}^{n}B_{g_1k}(X_i)\,\nu_k^2\right]\\
\simeq&E\left[(1-\xi_1)\sum_{i=1}^{n}\frac{1}{n^2g_2^2}\frac{\xi_iK^2\left(\frac{X_1-X_i}{g_2}\right)}{\pi^2(X_1)\,m^2(X_1)}\sum_{k=1}^{n}\frac{1}{ng_1}\frac{\xi_kK\left(\frac{X_i-X_k}{g_1}\right)}{\pi(X_i)\,m(X_i)}\nu_k^2\right]\\
=&\frac{1}{n^3g_2^2g_1}(n-1)\,K(0)\,E\left[(1-\xi_1)\,\xi_2\frac{\pi^{-1}(X_2)\,m^{-1}(X_2)}{\pi^2(X_1)\,m^2(X_1)}K^2\left(\frac{X_1-X_2}{g_2}\right)\nu_2^2\right]\\
&+\frac{1}{n^3g_2^2g_1}(n-1)(n-2)\,E\left[(1-\xi_1)\,\xi_2\xi_3\frac{\pi^{-1}(X_2)\,m^{-1}(X_2)}{\pi^2(X_1)\,m^2(X_1)}K\left(\frac{X_2-X_3}{g_1}\right)\right.\\
&\left.\times K^2\left(\frac{X_1-X_2}{g_2}\right)\nu_3^2\right]\\
=&\frac{1}{n^2g_2g_1}K(0)\,c_KE\left[\frac{1-\pi(X)}{\pi^2(X)\,m(X)}\left(\sigma^2(X)+\mu^2(X)\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{ng_2}c_K E\left[\frac{\pi\left(X\right)\left(1-\pi\left(X\right)\right)}{m\left(X\right)}\left(\sigma^2\left(X\right)+\mu^2\left(X\right)\right)\right] \\
&= o\left(1\right).
\end{aligned}
$$

The second expectation is

$$
\begin{aligned}
& \left(n-1\right)\left(n-2\right)E\left[\left(1-\xi_1\right)B_{g_2 2}\left(X_1\right)B_{g_2 3}\left(X_1\right)E\left(\nu_2^* \mid O\right)E\left(\nu_3^* \mid O\right)\right] \\
= & \left(n-1\right)\left(n-2\right)E\left[\left(1-\xi_1\right)\frac{1}{ng_2}\frac{\xi_2 K\left(\frac{X_1-X_2}{g2}\right)}{\pi\left(X_1\right)m\left(X_1\right)}\frac{1}{ng_2}\frac{\xi_3 K\left(\frac{X_1-X_3}{g2}\right)}{\pi\left(X_1\right)m\left(X_1\right)}\right. \\
& \left.\sum_{j=1}^{n}\frac{1}{ng_1}\frac{\xi_j K\left(\frac{X_j-X_2}{g1}\right)}{\pi\left(X_2\right)m\left(X_2\right)}\nu_j \sum_{k=1}^{n}\frac{1}{ng_1}\frac{\xi_k K\left(\frac{X_k-X_3}{g1}\right)}{\pi\left(X_3\right)m\left(X_3\right)}\nu_k\right] \\
= & \frac{1}{n^4 g_1^2 g_2^2}\left(n-1\right)\left(n-2\right)E\left[\left(1-\xi_1\right)\xi_2\xi_3\frac{\pi^{-1}\left(X_2\right)m^{-1}\left(X_2\right)\pi^{-1}\left(X_3\right)m^{-1}\left(X_3\right)}{\pi^2\left(X_1\right)m^2\left(X_1\right)}\right. \\
& \left.\sum_{j=1}^{n}\sum_{k=1}^{n}\xi_j\xi_k K\left(\frac{X_j-X_2}{g_1}\right)K\left(\frac{X_k-X_3}{g_1}\right)K\left(\frac{X_1-X_3}{g_2}\right)K\left(\frac{X_1-X_2}{g_2}\right)\nu_j\nu_k\right].
\end{aligned}
$$

There are different cases according to the values of j and k. The dominant term comes from the case $j \neq 1,2,3$ and $k \neq 1,2,3,j$, which gives

$$
E\left[\left(1-\pi\left(X\right)\right)\mu^2\left(X\right)\right]\left(1+o\left(g_1^2+g_2^2\right)\right).
$$

□

Lemma D.12 The integral $\iint K\left(u_1\right)K\left(u_1-\frac{g_2}{h}u_2\right)K\left(u_2\right)du_1du_2$ yields the following functions depending on the limit of the ratio of the bandwidths.

- If $\frac{g_2}{h}\to 0$,

  $$
  \iint K\left(u_1\right)K\left(u_1-\frac{g_2}{h}u_2\right)K\left(u_2\right)du_1du_2 = c_K + O\left(\frac{g_2^2}{h^2}\right),
  $$

  where $c_K = \int K^2\left(v\right)dv$.

- If $\frac{g_2}{h}\to C$,

  $$
  \iint K\left(u_1\right)K\left(u_1-\frac{g_2}{h}u_2\right)K\left(u_2\right)du_1du_2 = c_{K,C} + O\left(\frac{g_2^2}{h^2}\right),
  $$

  where $c_{K,C} = \iint K\left(u_1\right)K\left(u_2\right)K\left(u_1-Cu_2\right)du_1du_2$.

- If $\frac{g_2}{h} \to \infty$,

$$\iint K(u_1)\, K\left(u_1 - \frac{g_2}{h} u_2\right) K(u_2)\, du_1 du_2 = O\left(\frac{h}{g_2}\right) = o(1).$$

Proof. Starting with the first condition, if $\frac{g_2}{h} \to 0$, after applying Taylor expansion we get

$$\begin{aligned} &\iint K(u_1)\, K(u_2)\, du_1 du_2 K\left(u_1 - \frac{g_2}{h} u_2\right) = K(u_1) - \frac{g_2}{h} K'(u_2) + O\left(\frac{g_2^2}{h^2}\right) \\ = &\iint K^2(u_1)\, K(u_2)\, du_1 du_2 + O\left(\frac{g_2^2}{h^2}\right) = c_K + O\left(\frac{g_2^2}{h^2}\right). \end{aligned}$$

If $\frac{g_2}{h} \to C$ then

$$\begin{aligned} &\iint K(u_1)\, K\left(u_1 - \frac{g_2}{h} u_2 \pm C u_2\right) K(u_2)\, du_1 du_2 \\ = &\iint K(u_1)\, K\left(u_1 - C u_2 + \left(C - \frac{g_2}{h}\right) u_2\right) K(u_2)\, du_1 du_2 = K(u_1 - C u_2) \\ &+ O\left(C - \frac{g_2}{h}\right) \\ = &\iint K(u_1)\, K(u_1 - C u_2)\, K(u_2)\, du_1 du_2 + O\left(C - \frac{g_2}{h}\right) \\ = &\, c_{K,C} + O\left(C - \frac{g_2}{h}\right), \end{aligned}$$

where $c_{K,C} = \iint K(u_1)\, K(u_2)\, K(u_1 - C u_2)\, du_1 du_2$
If $\frac{g_2}{h} \to \infty$ then a change of variable $\omega_2 = \frac{g_2}{h} u_2$ yields

$$\begin{aligned} &\iint K(u_1)\, K\left(u_1 - \frac{g_2}{h} u_2\right) K(u_2)\, du_1 du_2 \\ = &\frac{h}{g_2} \iint K(u_1)\, K(u_1 - \omega_2)\, K\left(\frac{h}{g_2} \omega_2\right) du_1 d\omega_2 \\ = &\frac{h}{g_2} K(0) \iint K(u_1)\, K(u_1 - \omega_2)\, du_1 d\omega_2 + O\left(\frac{h^2}{g_2^2}\right) = O\left(\frac{h}{g_2}\right). \end{aligned}$$

□

Lemma D.13 The integral $\iiint K(u_1)\, K\left(u_1 - \frac{g_1}{h} u_3 - \frac{g_2}{h} u_2\right) K(u_2)\, K(u_3)\, du_1 du_2 du_3$ is:

- If $\frac{g_1}{h} \to 0$ and $\frac{g_2}{h} \to 0$ then

$$\iiint K(u_1)\, K\left(u_1 - \frac{g_1}{h} u_3 - \frac{g_2}{h} u_2\right) K(u_2)\, K(u_3)\, du_1 du_2 du_3 = c_K + o(1),$$

where $c_K = \int K^2(v)\, dv$.

- If one pilot bandwidth verifies $\frac{g_i}{h} \to 0$ and the other one $\frac{g_j}{h} \to C$ then

$$\iiint K(u_1) K\left(u_1 - \frac{g_1}{h}u_3 - \frac{g_2}{h}u_2\right) K(u_2) K(u_3)\, du_1 du_2 du_3 = c_{K,C} + o(1),$$

where $c_{K,C} = \iint K(u_1) K(u_2) K(u_1 - Cu_2)\, du_1 du_2$.

- If $\frac{g_1}{h} \to C_1$ and $\frac{g_2}{h} \to C_2$ then

$$\iiint K(u_1) K\left(u_1 - \frac{g_1}{h}u_3 - \frac{g_2}{h}u_2\right) K(u_2) K(u_3)\, du_1 du_2 du_3 = c_{K,C_1,C_2} + o(1),$$

where $c_{K,C_1,C_2} = \iint K(u_1) K(u_2) K(u_1 - C_1 u_3 - C_2 u_2)\, du_1 du_2$.

- If $\frac{g_1}{h} \to \infty$ or $\frac{g_2}{h} \to \infty$ then

$$\iiint K(u_1) K\left(u_1 - \frac{g_1}{h}u_3 - \frac{g_2}{h}u_2\right) K(u_2) K(u_3)\, du_1 du_2 du_3 = o(1).$$

Proof. If $\frac{g_1}{h} \to 0$ and $\frac{g_2}{h} \to 0$, we apply a Taylor expansion

$$\begin{aligned}
&\iiint K(u_1) K\left(u_1 - \frac{g_1}{h}u_3 - \frac{g_2}{h}u_2\right) K(u_2) K(u_3)\, du_1 du_2 du_3 \\
&= \iiint K(u_1) K(u_1) - \frac{g_1}{h}u_3 K'(u_1) - \frac{g_2}{h}u_2 K'(u_1) \\
&\quad + O\left(\frac{g_1^2}{h} + \frac{g_1}{h}^2\right) K(u_2) K(u_3)\, du_1 du_2 du_3 \\
&= c_K + O\left(\frac{g_1^2}{h} + \frac{g_1^2}{h}\right).
\end{aligned}$$

If $\frac{g_1}{h} \to 0$ and $\frac{g_2}{h} \to C_2$ then

$$\begin{aligned}
&\iiint K(u_1) K\left(u_1 - \frac{g_1}{h}u_3 - \frac{g_2}{h}u_2 \pm C_2 u_2\right) K(u_2) K(u_3)\, du_1 du_2 du_3 \\
&= \iiint K(u_1) K\left(u_1 - \frac{g_1}{h}u_3 - C_2 u_2 + \left(C_2 - \frac{g_2}{h}\right) u_2\right) K(u_2) K(u_3)\, du_1 du_2 du_3 \\
&= \iint K(u_1) K(u_2) K(u_1 - C_2 u_2)\, du_1 du_2 + O\left(\frac{g_1}{h}\right) + O\left(C_2 - \frac{g_2}{h}\right) \\
&= c_{K,C_1} + O\left(\frac{g_1}{h}\right) + O\left(C_2 - \frac{g_2}{h}\right)
\end{aligned}$$

where $c_{K,C} = \iint K(u_1) K(u_2) K(u_1 - Cu_2)\, du_1 du_2$.

If $\frac{g_1}{h} \to C_1$ and $\frac{g_2}{h} \to C_2$ then

$$\begin{aligned}
&\iiint K(u_1) K\left(u_1 - \frac{g_1}{h}u_3 - \frac{g_2}{h}u_2 \pm C_1 u_3 \pm C_2 u_2\right) K(u_2) K(u_3)\, du_1 du_2 du_3 \\
&= \iiint K(u_1) K\left(u_1 - C_1 u_3 - C_2 u_2 + \left(C_1 - \frac{g_1}{h}\right) u_3 + \left(C_2 - \frac{g_2}{h}\right) u_2\right) \\
&\quad \times K(u_2) K(u_3)\, du_1 du_2 du_3 \\
&= \iint K(u_1) K(u_2) K(u_1 - C_1 u_3 - C_2 u_2)\, du_1 du_2 + O\left(C_1 - \frac{g_1}{h}\right) + O\left(C_2 - \frac{g_2}{h}\right) \\
&= c_{K,C_1,C_2} + O\left(C_1 - \frac{g_1}{h}\right) + O\left(C_2 - \frac{g_2}{h}\right)
\end{aligned}$$

where $c_{K,C_1,C_2} = \iint K(u_1) K(u_2) K(u_1 - C_1 u_3 - C_2 u_2)\, du_1 du_2$.
If $\frac{g_1}{h} \to 0$ and $\frac{g_2}{h} \to \infty$ then a change of variable $\omega_2 = \frac{g_2}{h}u_2$ and a Taylor expansion yield

$$\frac{h}{g_2} \iiint K(u_1) K\left(u_1 - \frac{g_1}{h}u_3 - \omega_2\right) K\left(\frac{h}{g_2}\omega_2\right) K(u_3)\, du_1 du_2 du_3 = O\left(\frac{h}{g_2}\right) = o(1).$$

Using the same steps, if $\frac{g_1}{h} \to \infty$ then

$$\iiint K(u_1) K\left(u_1 - \frac{g_1}{h}u_3 - \frac{g_2}{h}u_2\right) K(u_2) K(u_3)\, du_1 du_2 du_3 = o(1).$$

□

Lemma D.14 The integral
$\iiint K\left(u_2 + \frac{g_2}{h}(u_2 + u_4) + \frac{g_1}{h}u_3\right) K(u_2) K(u_3) K(u_4)\, du_2 du_3 du_4$ becomes

- If $\frac{g_1}{h} \to 0$ and $\frac{g_2}{h} \to 0$ then

$$\iiint K\left(u_2 + \frac{g_2}{h}(u_2 + u_4) + \frac{g_1}{h}u_3\right) K(u_2) K(u_3) K(u_4)\, du_2 du_3 du_4 = c_K + o(1),$$

  where $c_K = \int K^2(v)\, dv$.

- If $\frac{g_1}{h} \to 0$ and $\frac{g_2}{h} \to C_2$ then

$$\iiint K\left(u_2 + \frac{g_2}{h}(u_2 + u_4) + \frac{g_1}{h}u_3\right) K(u_2) K(u_3) K(u_4)\, du_2 du_3 du_4 = d_{K,C_2} + o(1),$$

  where $d_{K,C_2} = \iiint K(u_2 + C_2(u_2 + u_4)) K(u_2) K(u_4)\, du_2 du_4$.

- If $\frac{g_1}{h} \to C_1$ and $\frac{g_2}{h} \to 0$ then

$$\iiint K\left(u_2 + \frac{g_2}{h}(u_2 + u_4) + \frac{g_1}{h}u_3\right) K(u_2) K(u_3) K(u_4)\, du_2 du_3 du_4 = c_{K,C_1} + o(1),$$

where $c_{K,C} = \iiint K(u_2 + Cu_3) K(u_2) K(u_3) du_2 du_3$ or equivalently $c_{K,C} = \iint K(u_1) K(u_2) K(u_1 - Cu_2) du_1 du_2$.

- If $\frac{g_1}{h} \to C_1$ and $\frac{g_2}{h} \to C_2$ then

  $$\iiint K\left(u_2 + \frac{g_2}{h}(u_2 + u_4) + \frac{g_1}{h}u_3\right) K(u_2) K(u_3) K(u_4) du_2 du_3 du_4 = d_{K,C_1,C_2} + o(1),$$

  where $d_{K,C_1,C_2} = \iiint K(u_2 + C_2(u_2 + u_4) + C_1 u_3) K(u_2) K(u_3) K(u_4) du_2 du_3 du_4$.

- If $\frac{g_1}{h} \to \infty$ or $\frac{g_2}{h} \to \infty$ then

  $$\iiint K\left(u_2 + \frac{g_2}{h}(u_2 + u_4) + \frac{g_1}{h}u_3\right) K(u_2) K(u_3) K(u_4) du_2 du_3 du_4 = o(1).$$

Appendix E

Resumen en español

Los métodos clásicos de análisis de supervivencia asumen que, si el período de seguimiento es suficientemente largo y no hay censura, todos los individuos experimentarán el suceso de interés. Sin embargo, hay muchos ejemplos en los que hay evidencia de pacientes que nunca experimentarán el suceso, llamados supervivientes a largo plazo o curados. En este caso se deberán usar en su lugar los modelos de curación. Bajo censura, estos modelos asumen que no es posible determinar si un individuo censurado experimentará en el futuro el evento o no. Solo se puede saber que los sucesos observados corresponden a pacientes no curados, pero no es posible distinguir ningún individuo curado. En consecuencia, el indicador de cura se modeliza como una variable latente. Sin embargo esto no es necesariamente cierto en muchos casos, en los que algunos individuos censurados se pueden identificar como curados, basándose por ejemplo en un test diagnósticados o si el tiempo de vida supera un determinado umbral.

Los modelos de curación de tipo mixtura se han estimado normalmente usando técnicas paramétricas o semiparamétricas. Recientemente se ha propuesto un enfoque completamente no paramétrico, cuando se desconoce completamente si un sujeto censurado está curado o no. Esta tesis propone una extensión a los modelos no paramétricos de curación de tipo mixtura, en la que se incorporaría la información adicional disponible sobre el estado de cura. Se proponen estimadores no paramétricos de las principales funciones, así como un sencillo método para comprobar la validez del modelo. Los métodos se han aplicado a tres bases de datos médicos: una relacionada con pacientes de sarcoma, otra de pacientes con cáncer de pecho, y una tercera sobre las duraciones de estancia en planta y UCI de pacientes COVID-19 en Galicia durante la primera ola de la pandemia en 2020.

1. Introducción

El primer capítulo de la tesis está dedicado a introducir el contexto en el que se desarrolla la tesis: los modelos de curación de tipo mixtura. La Sección 1.1 comienza con una revisión de los principales estimadores clásicos de la función de supervivencia bajo censura (Kaplan and Meier, 1958; Beran, 1981).

En la Sección 1.2 se presenta una descripción detallada de los modelos de curación, comenzando con la notación y el problema de la identificabilidad, que surge principalmente debido a la falta de información en la cola derecha de la distribución por la censura. Los modelos de curación se pueden clasificar entre los de tipo mixtura y no mixtura. Esta tesis se centra en los modelos de curación de tipo mixtura (MCM), que clasifican a los individuos en dos grupos: los que tarde o temprano experimentarán el suceso de interés (susceptibles) y los que no (curados). De esta manera, estos modelos permiten estimar la probabilidad de experimentar el suceso (*incidencia*), o su complementario la probabilidad de no experimentarlo (*cura*), así como la función de supervivencia de los individuos susceptibles (*latencia*). Una de las principales ventajas de los modelos de curación de tipo mixtura es que permiten ajustar de manera independiente el efecto que una covariable tendrá en la incidencia y en la latencia, así como considerar el hecho de que puedan ser distintas covariables las que pueden tener influencia en los pacientes curados o en los susceptibles.

A continuación, se hace una breve revisión de los métodos clásicos de estimación en los modelos de curación, principalmente desde un punto de vista no paramétrico (Maller and Zhou, 1992; Xu and Peng, 2014; López-Cheda et al., 2017a,b). A lo largo de la tesis, se hará uso del método bootstrap para la selección del parámetro ventanta. En la Sección 1.3 se revisa los distintos métodos de remuestreo bajo censura en un contexto incondicional (Efron, 1981), en presencia de covariables continuas (Li and Datta, 2001) y en presencia de individuos curados (López-Cheda et al., 2017a,b). Por su parte, en la Sección 1.4 se revisan los principales métodos de selección de la ventana. Finalmente, en la Sección 1.5 se aborda el contexto particular de la tesis, que es un modelo de curación tipo mixtura en la que algunos individuos censurados se pueden identificar como curados, cuyos tiempos observados son aleatorios. En concreto, se introduce la notación que se va a usar a lo largo de la memoria.

El capítulo termina con la presentación de los ejemplos de datos reales que motivan la metodología presentada en la tesis, y que se van a usar en los siguientes capítulos a modo ilustrativo. El primer ejemplo hace referencia a pacientes diagnosticados con sarcoma, donde el evento de interés es la muerte por complicaciones debido al sarcoma. Cuando el tumor se elimina quirúrgicamente y el paciente permanece libre de la enfermedad durante al menos 5 años, se puede asumir que el sarcoma ha remitido (Choy, 2014).

Los pacientes que cumplan ambos requisitos se pueden asumir, desde un punto de vista estadístico, como *curados de muerte por sarcoma.* Este base de datos se ha utilizado en el capítulo 2 para ilustrar el estimador propuesto para la función de supervivencia.

El segundo ejemplo es una base de datos de pacientes con cáncer de pecho, y el suceso de interés es, de nuevo, la muerte debido al cáncer. Por lo tanto, el tiempo observado de una paciente se considera censurado si no se observó el suceso de interés durante el período de seguimiento, es decir, si la paciente estaba viva al finalizar el estudio, estaba libre de la enfermedad (independientemente de que hubiera fallecido o no), o hubo una pérdida de seguimiento. Si una paciente está libre de la enfermedad durante al menos 10 años se pueden considerar como que no va a fallecer por culpa del cáncer (Barnadas et al., 2018), desde un punto de vista estadístico, está *curada de fallecer debido al cáncer.* En esta base de datos, el objetivo será estimar la probabilidad de no fallecer debido al cáncer, usando la metodología del capítulo 3.

La tercera base de datos está relacionada con los pacientes diagnosticados de COVID-19 durante la primera ola de la pandemia en Galicia, entre los meses de marzo y mayo de 2020. El enfoque en esta base de datos no está tan relacionado con un objetivo médico como los anteriores, sino más bien con la gestión hospitalaria. Durante las primeras semanas de pandemia, resultó de extrema necesidad planificar y estimar adecuadamente la ocupación de camas tanto de planta como de UCI, con el fin de evitar sobrecargas del sistema hospitalario gallego. Para ello, resultaba básico modelizar los tiempos de estancia de estos enfermos en los centros hospitalarios. En concreto, se deseaba saber cuál era la probabilidad de que un paciente ingresado en planta requiriese finalmente ingresar en UCI, y cuánto tardaban estos pacientes en ingresar en UCI. Usando un lenguaje de análisis de supervivencia, el suceso de interés es la entrada de un paciente COVID en UCI desde planta. No todos los pacientes que pasaron por planta necesitaron ingresar en UCI, por lo que era muy razonable ajustar un modelo de curación para modelizar los tiempos de estancia hasta entrada en UCI. Aquellos pacientes todavía ingresados en planta al finalizar el estudio eran datos censurados, para los cuales no se sabía si acabarían necesitando UCI o no. Sin embargo, todos los pacientes fallecidos en planta y todos aquellos dados de alta antes de entrar en UCI también son datos censurados, pero a diferencia de los anteriores, de estos pacientes sí se sabe que no entrarán nunca en UCI (*curados del suceso entrada en UCI*). Con esta base de datos se han usado los métodos del capítulo 4 para estimar la probabilidad de que un paciente COVID ingresado en planta requiera tratamiento en UCI, así como los estimadores del capítulo 5 para modelizar la distribución de los tiempos de estancia en planta hasta ingreso en UCI de los pacientes que finalmente sí entraron en UCI.

2. Estimador límite-producto generalizado de la función de supervivencia

En este capítulo, después de una breve introducción y de presentar una serie de definiciones y de hipótesis en la Sección 2.2, en la Sección 2.3 de este capítulo se presenta la primera aportación de esta tesis, un estimador no paramétrico límite-producto generalizado de la función de supervivencia condicional en el MCM cuando el estado de cura es parcialmente conocido, es decir, cuando algunos individuos censurados son identificados como individuos que no experimentarán el suceso de interés. Este estimador es la extensión del estimador de Beran (1981) al caso de cura parcialmente conocida. De hecho, el estimador propuesto se puede ver como el estimador de Beran (1981) calculado con los tiempos de vida observados, donde los tiempos de los individuos identificados como curados se sustituyen por valores arbitrariamente grandes, incluso infinito. Además, se da la expresión que este estimador toma de forma incondicional cuando no hay ninguna covariable, y se propone el correspondiente estimador de la distribución de censura en este contexto.

La Sección 2.3.1 muestra las principales propiedades asintóticas de dicho estimador, en concreto, se obtiene una representación casi segura como suma de términos iid, de la cual se deriva la consistencia fuerte y la normalidad asintótica del estimador. A partir de las expresiones asintóticas del sesgo y la varianza del estimador propuesto, se demuestra en la Sección 2.3.2 que la incorporación del estado de cura en la estimación produce una reducción en el término dominante de la varianza con respecto a estimador de Beran.

A partir de los estimadores propuestos para la función de supervivencia de los tiempos de vida, de la variable de censura así como de los tiempos observados, en la Sección 2.3.3 se proponen dos métodos de remuestreo bootstrap, llamados *simple weighted bootstrap* y *obvious bootstrap* siguiendo la notación en Li and Datta (2001), y se prueba que ambos métodos de remuestreo son equivalentes.

En la Sección 2.3.4 se introduce un método de selección del parámetro ventana basado en el bootstrap. A continuación, en el estudio de simulación de la Sección 2.4 se compara el estimador propuesto de la función de supervivencia con el estimador que ignora la cura conocida (Beran, 1981) y el estimador semiparamétrico de Bernhardt (2016), que asume una función logística para la probabilidad de cura y ajusta un model AFT para la latencia. Los resultados muestran que, si el parámetro ventana se elige adecuadamente, nuestro estimador funciona mejor que los otros estimadores considerados para un amplio rango de valores de la covariable.

Por último, en la Sección 2.5 el estimador se aplica a un conjunto de datos real que estudia la supervivencia de los pacientes con sarcoma. En este ejemplo, no se aprecian grandes diferencias entre el estimador propuesto y el estimador de Beran, puesto que los tiempos observados de los pacientes considerados curados de la muerte por sarcoma son valores muy altos y, por tanto, la mejora en la estimación de la función de supervivencia que supone tener en cuenta la cura conocida es limitada.

Los resultados incluidos en este capítulo se han publicado en Safari et al. (2021).

3. Estimador kernel de la probabilidad de cura

En este capítulo, en la Sección 3.1 se introduce el estimador propuesto para estimar la probabilidad de cura, basado en el estimator de la función de supervivencia presentado en el capítulo anterior. Se trata de la extensión del estimador propuesto por Xu and Peng (2014) al contexto en el que algunos individuos censurados se pueden identificar como curados del evento de interés. Se estudian las propiedades asintóticas de este estimador, tales como la representación iid, la consistencia fuerte y se prueba que su distribución es asintóticamente normal. Además, se dan las expresiones de los términos dominantes del sesgo y la varianza, y se muestra cuál es el efecto en dichos términos de ignorar el hecho de que algunos individuos censurados son en realidad curados. Finalmente, se propone un método para seleccionar el parámetro ventana usando un procedimiento bootstrap.

En la Sección 3.3 se presentan dos estimadores para la probabilidad de cura, también no paramétricos de tipo kernel, que no han sido estudiados previamente en la literatura. El primero de ellos, en la Sección 3.3.1, está basado en el modelo de riesgos competitivos. Siguiendo la propuesta de Betensky and Schoenfeld (2001) para un contexto incondicional, se puede asumir que el modelo de curación tipo mixtura con cura parcialmente conocida en la que los tiempos de curación son aleatorios es un modelo de riesgos competitivos en los que hay dos únicos tipos de riesgos, el evento de interés y la cura. En este caso, la probabilidad de cura no es más que el límite de la función de incidencia acumulada de la cura, o equivalentemente, uno menos el límite de la función de incidencia acumulada del evento de interés.

La segunda alternativa propuesta para estimar la probabilidad de cura se presenta en la Sección 3.3.2. La idea principal detrás de este estimador es que, debido a que el indicador de cura es una variable binaria, la probabilidad de cura se puede escribir también como la esperanza condicional del indicador de cura. De esta forma, se pueden aplicar métodos de regresión no paramétrica para estimar dicha esperanza condicional. Sin embargo, en presencia de censura, no se conoce el indicador de cura para todos los individuos, puesto que para muchos individuos censurados se desconoce si finalmente experimentarán el suceso de interés o no. La aplicación de métodos de regresión en este contexto requiere manejar los datos faltantes en la variable respuesta (indicador de cura). Hay en la literatura distintas propuestas para ajustes de regresión con valores perdidos en la variable respuesta (Hsu et al., 2016; Verhasselt et al., 2019; Vakulenko-Lagun et al., 2020). Aerts et al. (2002) propusieron un método de imputación múltiple para la estimación de la esperanza incondicional. En esta sección se extienden esas ideas al caso con una covariable continua, en el que la esperanza condicional se estima esperanzante el estimador de Nadaraya-Watson (NW) aplicado a los datos en los que la respuesta faltante se ha sustituido previamente por los valores imputados. Este estimador es consistente bajo la hipótesis *strongly ignorable missing at random* (siMAR), es decir, la probabilidad de observar el valor del indicador de cura depende de la covariable X pero no del propio indicador de cura. En el modelo de mixtura con cura parcialmente conocida, esta hipótesis no es cierta si el tiempo de vida Y y la variable de censura C son condicionalmente independientes, puesto que la probabilidad de observar el indicador de cura es mayor en los individuos susceptibles que en los curados, y por tanto, la probabilidad de observar el valor del indicador de cura depende del propio indicador de cura. Sin embargo, cuanto mayor sea el porcentaje de sujetos curados que se observan, más débil es la relación entre la probabilidad de observar el valor del indicador de cura y el propio indicador de cura, y por tanto, más próxima está la hipótesis siMAR de ser cierta.

En la Sección 3.4 se lleva a cabo un estudio de simulación para evaluar el comportamiento del estimador de la probabilidad de cura propuesto, basado en el modelo de mixtura con cura, así como el método de selección de la ventana, en comparación con los estimadores alternativos propuestos, tanto los basados en un modelo de riesgos competitivos como los que se obtienen con un ajuste de regresión, previa imputación de los indicadores de cura desconocidos. Además, a efectos de comparación, se consideran también el estimador que ignora la cura conocida (Xu and Peng, 2014) y el estimador semiparamétrico de Bernhardt (2016), que sí incorpora la información de los individuos curados, y que estima la probabilidad de cura esperanzante un algoritmo EM, ajustando una regresión logística.

La distribución de los tiempos de los individuos susceptibles se simularon usando una distribución exponencial truncada. Por su parte, la distribución de censura es impropia, en la que $C = \infty$ con probabilidad $\pi(x)$ y en caso contrario C se simuló de una distribución Weibull. Se simularon 6 escenarios distintos en función de la expresión de la probabilidad de cura, y suponiendo que el porcentaje de individuos curados que son observados era $\pi(x) = 0.2$ y $\pi(x) = 0.8$.

Los primeros resultados muestran el error cuadrático medio (MSE) de los estimadores. En el caso de los estimadores suavizados, se muestra el MSE con la correspondiente ventana óptima, seleccionada de una rejilla de 21 ventanas equiespaciadas entre 1.5 y 100. La principal conclusión es que el estimador propuesto, basado en el modelo de mixtura con cura, es muy competitivo en todos los escenarios simulados. Este estimador mejora claramente al estimador que ignora la información de la cura, da resultados similares o ligeramente mejores al estimador basado en el modelo de riesgos competitivos, también mejora claramente al basado en modelos de regresión cuando el porcentaje de pérdida de información es alto ($\pi(x)$ bajo), y al estimador semiparamétrico cuando la probabilidad de cura no se ajusta al modelo paramétrico logístico. En el escenario 1, en el que la expresión de la probabilidad de cura es logística, el estimador semiparamétrico se comporta muy bien, aunque sus resultados son más pobres en los demás escenarios. En los resultados relacionados con la eficiencia del selector bootstrap de la ventana, podemos apreciar que la ventana bootstrap se aproxima correctamente a la ventana óptima teórica.

A continuación, en la Sección 3.5, todos los estimadores de la probabilidad de cura estudiados en el presente capítulo se aplican a dos bases de datos reales. La primera de ellas, relacionada con pacientes con cáncer de pecho, se analiza en la Sección 3.5.1. El objetivo es estimar la probabilidad de no morir de cáncer de mama dependiendo del estadío de la enfermedad, el número de ganglios linfáticos positivos, si la paciente es menopáusica o no, y la edad en el momento del diagnóstico. Aunque el estimador semiparamétrico sugiere que la probabilidad de no morir por cáncer de mama aumenta con la edad, los demás estimadores que son no paramétricos indican que un ajuste logístico podría no ser adecuado, indicando que la probabilidad aumenta solo en mujeres jóvenes y de mediana edad, mientras que la edad no tiene efecto en las mujeres más mayores. Por último, el estimador MI-NW proporciona estimaciones sensiblemente menores que los demás estimadores, debido principalmente a que hay pocas mujeres que se puedan identificar como pacientes curadas de fallecer por la enfermedad, y por tanto el porcentaje de pérdida en el indicador de cura es muy alto.

El segundo ejemplo con datos reales, en el que se analizan pacientes con COVID-19, se muestra en la Sección 3.5.2. Un conocimiento preciso de la duración de la hospitalización y la predicción de la probabilidad de que un paciente hospitalizado requiera una cama en la UCI, son clave para comprender la demanda hospitalaria de camas y crucial para la toma de decisiones y una planificación adecuada. En este análisis, el tiempo de interés es la duración de la estancia en planta de los pacientes hasta el ingreso en la UCI, y el objetivo es estimar la probabilidad de ingreso en la UCI en función de la edad y el sexo como covariables de interés. Aunque el estimador semiparamétrico sugiere un efecto decreciente uniforme de la edad sobre la probabilidad de ingreso en la UCI, los estimadores no paramétricos indican que el ajuste logístico no parece razonable ya que las curvas se caracterizan por una probabilidad de ingreso en la UCI constante o ligeramente creciente para los pacientes más jóvenes (menores de 55 años), un fuerte aumento de la probabilidad para pacientes de mediana edad (de 55 a 69 años) y una disminución para pacientes de edad avanzada (70 años o más). El estimador XP sobreestima la probabilidad de admisión en UCI, debido a que no tiene en cuenta que muchos pacientes nunca necesitarán UCI, mientras que el estimador MI-NW infraestima dicha probabilidad para los pacientes más jóvenes, puesto que hay un número muy bajo de pacientes jóvenes cuya entrada en UCI ha sido observada, por lo que hay un porcentaje muy alto de pérdida en el indicador de cura para dichos pacientes.

Los resultados incluidos en este capítulo se pueden encontrar en Safari et al. (2022a).

4. Estimador no paramétrico de la latencia

La tercera contribución de la tesis, incluida en este capítulo, aborda el problema de la estimación de la función de latencia. El estimador de latencia propuesto extiende el estimador no paramétrico estudiado en López-Cheda et al. (2017b) para el caso en que el estado de curación se observe para algunos individuos censurados. En la Sección 4.2 se da la expresión del estimador propuesto, tanto en un contexto condicional para una covariable continua, como para el caso incondicional. A diferencia del estimador propuesto por López-Cheda et al. (2017b) que depende de una única ventana, el estimador que se presenta en esta tesis depende de dos ventanas distintas, lo que aporta más flexibilidad a la estimación. Los resultados asintóticos, tales como la representación iid y la distribución asintóticamente normal, se muestran en la Sección 4.2.1. En la Sección 4.2.3 se proporciona un método de selección de las ventanas basado en el bootstrap.

En la Sección 4.3, un estudio de simulación muestra los resultados del estimador propuesto, en comparación con el estimador que ignora la cura conocida (López-Cheda et al., 2017b), el estimador semiparamétrico que ajusta un modelo AFT (Bernhardt, 2016), el efecto de usar dos ventanas en la estimación sobre el estimador con una única ventana, así como el comportamiento en la práctica del selector de la ventana tipo bootstrap. Tal como se esperaba, el estimador semiparamétrico solo da buenos resultados en los modelos para los cuales el modelo AFT es adecuado. Claramente, el estimador con dos ventanas mejora al que se calcula con una única ventana. Finalmente, el estimador propuesto es más eficiente que el estimador que ignora la cura conocida, especialmente en los escenarios simulados en los que la cura se observa con una alta probabilidad.

El capítulo termina con una aplicación a la base de datos real de pacientes COVID-19 en la Sección 4.4, el estimador se aplica para estudiar la duración de la estancia hospitalaria de los pacientes con COVID-19 que requieren cuidados intensivos. Las estimaciones del estimador propuesto y el estimador semiparamétrico son similares, lo que sugiere que el ajuste AFT podría ser adecuado. Sin embargo, las estimaciones con el método de López-Cheda et al. (2017b) que no tiene en cuenta la cura conocida aparentan demasiado optimistas. Para finalizar, se incluye una breve discusion en la Sección 4.5.

Los resultados que figuran en este capítulo se incluyen en Safari et al. (2022b)

5. Análisis de la independencia entre el tiempo de vida y la censura

Una característica importante de los estimadores propuestos en esta tesis, al igual que la mayoría de los procedimientos en análisis de supervivencia, es que son consistentes solamente bajo el supuesto de independencia condicional del tiempo de vida y la variable de censura. Este supuesto implica que el mecanismo que induce a la censura es totalmente ajeno al evento de interés. La independencia entre Y y C es bastante natural y se puede asumir con mucha frecuencia en la mayoría de los casos. Esta independencia entre el tiempo de vida y el tiempo de censura es una hipótesis crucial para hacer inferencias insesgadas en análisis de supervivencia. A pesar de esto, casi nunca se comprueba específicamente en la práctica. El problema es que no se puede testar cuando los datos incluyen solo el tiempo observado, posiblemente censurado, y un indicador de censura. Al no observar nunca el valor de Y y C de forma simultánea, siempre se puede obtener con los datos observados una expresión para las distribuciones de Y y C que cumplen la hipótesis de independencia (Tsiatis, 1975). Así que no hay ninguna prueba estadística formal para comprobar si el tiempo de censura es independiente del tiempo de vida sin asumir más hipótesis para la distribución conjunta de Y y C.

En este capítulo se presenta un procedimiento simple no paramétrico para evaluar hasta qué punto es plausible el supuesto de independencia. Este método se basa en el hecho de que, bajo la hipótesis de independencia entre Y y C, el estimador propuesto para la probabilidad de cura es insesgado y consistente, mientras que el estimador MI-NW sería sesgado e inconsistente puesto que no se verifica la hipótesis siMAR. Como resultado, cuando el porcentaje de censura es alto, ambos estimadores darían resultados muy diferentes. Valores altos de la diferencia entre ambos estimadores serían consistentes con la hipótesis de independencia, mientras que valores similares evidenciarían que la hipótesis no es verosímil.

En la Sección 5.1 se hace una revisión al problema de falta de independencia entre Y y C, en la Sección 5.2 se motiva por qué la diferencia entre las estimaciones de la probabilidad de cura dadas por el estimador propuesto y el estimador MI-NW se puede usar para medir la plausibilidad de la hipótesis de independencia, mientras en la Sección 5.2.1 se muestran los resultados de un estudio sobre la sensibilidad del procedimiento a distintos grados de dependencia y distintos niveles de observación de la cura.

Conclusiones y trabajo futuro

En este último capítulo de la memoria de tesis, se hace un resumen de los principales resultados obtenidos en los capítulos anteriores. A continuación, se expone una serie de líneas en las que podría avanzar la investigación en el futuro.

La primera línea de trabajo futuro hace referencia al capítulo 5, se llevará a cabo un estudio de simulación extenso en el que se evaluará el test estadístico de contraste, se aproximará su distribución bajo la hipótesis nula esperanzante bootstrap, y se propondrá un selector para el parámetro ventana.

Además, los métodos propuestos están diseñados para el caso más general en el que los tiempos están sujetos a censura aleatoria por la derecha. No obstante, las observaciones pueden sufrir otros tipos de censura, o incluso truncamiento. La metodología presentada en esta memoria se adaptarán a este tipo de datos más complejos.

En muchos casos, el interés reside más allá de una sola covariable continua X, cuando se dispone de múltiples covariables $\boldsymbol{X}$, un vector de covariables de naturaleza posiblemente mixta, es decir, con componentes discretas, categóricas y/o variables continuas. Para evitar el problema de la maldición de la multidimensionalidad, que surge en la estimación no paramétrica cuando el número de covariables no va acompañado con un aumento suficiente en el tamaño muestral, nos centraremos en dos enfoques distintos. El primero se basa en el uso de funciones núcleo multivariantes definidas por el producto de las correspondientes funciones núcleo univariantes. El segundo consiste en utilizar un modelo *single index*, en el que el efecto de todas las covariables se resuma en un único índice unidimensional, dado por una función de una combinación lineal de las covariables, es decir, $g(\beta^T \boldsymbol{X})$.

Otra línea de investigación futura será los contrastes de hipótesis asociados a selección de covariables y a los test de significación, así como la introducción de la cura parcialmente conocida en los modelos de riesgos competitivos y, más en general, en los modelos multiestado.

Finalmente, el paquete `npcure` de `R` de López-Cheda et al. (2021) permite obtener las estimaciones no paramétricas de la función de supervivencia, latencia y la probabilidad de cura propuestos por López-Cheda et al. (2017a,b, 2020) que ignora la cura conocida, incluido el estimador de Beran. El paquete no incluye la situación en la que el estado de curado es parcialmente conocido. Los métodos de estimación propuestos en esta tesis se incluirán en futuras versiones del paquete.

References

Aerts, M., Claeskens, G., Hens, N., and Molenberghs, G. (2002). Local multiple imputation. *Biometrika*, 89(2):375–388.

Akritas, M. G. (1986). Bootstrapping the Kaplan-Meier estimator. *Journal of the American Statistical Association*, 81(396):1032–1038.

Amico, M. (2018). *Cure Models in Survival Analysis: From Modelling to Prediction Assessment of the Cure Fraction.* PhD thesis, Université Catholique de Louvain & Katholieke Universiteit Leuven. `https://dial.uclouvain.be/pr/boreal/en/object/boreal%3A208410` [Online: accessed 06-April-2022].

Amico, M. and Van Keilegom, I. (2018). Cure models in survival analysis. *Annual Review of Statistics and Its Application*, 5:311–342.

Amico, M., Van Keilegom, I., and Legrand, C. (2019). The single-index/Cox mixture cure model. *Biometrics*, 75(2):452–462.

Anderson, N. H., Hall, P., and Titterington, D. M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *Journal of Multivariate Analysis*, 50(1):41–54.

Andrei, Y. Y. and Asselain, B. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications.* World Scientific.

Andridge, R. R. and Little, R. J. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review*, 78(1):40–64.

Barnadas, A., Algara, M., Cordoba, O., Casas, A., González, M., Marzo, M., Montero, A., Muñoz, M., Ruiz, A., and Santolaya, F. (2018). Recommendations for the follow-up care of female breast cancer survivors: a guideline of the Spanish Society of Medical Oncology (SEOM), Spanish Society of General Medicine (SEMERGEN), Spanish Society for Family and Community Medicine (SEMFYC), Spanish Society for General and Family Physicians (SEMG), Spanish Society of Obstetrics and Gynecology (SEGO), Spanish Society of Radiation Oncology (SEOR), Spanish Society of Senology and Breast Pathology (SESPM), and Spanish Society of Cardiology (SEC). *Clinical and Translational Oncology*, 20(6):687–694.

Beesley, L. J. and Taylor, J. M. (2019). EM algorithms for fitting multistate cure models. *Biostatistics*, 20(3):416–432.

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

Bernhardt, P. W. (2016). A flexible cure rate model with dependent censoring and a known cure threshold. *Statistics in Medicine*, 35(25):4607–4623.

Betensky, R. A. and Schoenfeld, D. A. (2001). Nonparametric estimation in a cure model with random cure times. *Biometrics*, 57(1):282–286.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B*, 11(1):15–53.

Bowman, A., Hall, P., and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika*, 85(4):799–808.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product-limit estimates under random censorship. *The Annals of Statistics*, 2(3):437–453.

Cao, R. and González-Manteiga, W. (1993). Bootstrap methods in regression smoothing. *Journal of Nonparametric Statistics*, 2:379–388.

Cao, R. and Van Keilegom, I. (2006). Empirical likelihood tests for two-sample problems via nonparametric density estimation. *Canadian Journal of Statistics*, 34(1):61–77.

Carbonnaux, M., Brahmi, M., Schiffler, C., Meeus, P., Sunyach, M.-P., Bouhamama, A., Karanian, M., Tirode, F., Pissaloux, D., Vaz, G., et al. (2019). Very long-term survivors among patients with metastatic soft tissue sarcoma. *Cancer Medicine*, 8(4):1368–1378.

Carpenter, J. and Kenward, M. (2012). *Multiple Imputation and its Application.* John Wiley & Sons, New York, NY.

Chen, M., Ibrahim, J. G., and Lipsitz, S. R. (2002). Bayesian methods for missing covariates in cure rate models. *Lifetime Data Analysis*, 8(2):117–146.

Chen, M., Ibrahim, J. G., and Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94(447):909–919.

Chen, T. and Du, P. (2018). Promotion time cure rate model with nonparametric form of covariate effects. *Statistics in Medicine*, 37(10):1625–1635.

Cheng, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association*, 89(425):81–87.

Chown, J., Heuchenne, C., and Van Keilegom, I. (2020). The nonparametric location-scale mixture cure model. *TEST*, 29(4):1008–1028.

Choy, E. (2014). Sarcoma after 5 years of progression-free survival: Lessons from the French sarcoma group. *Cancer*, 120(19):2942–2943.

Colomer, R., Aranda-López, I., Albanell, J., García-Caballero, T., Ciruelos, E., López-García, M., Cortés, J., Rojo, F., Martín, M., and Palacios-Calvo, J. (2018). Biomarkers in breast cancer: A consensus statement by the Spanish Society of Medical Oncology and the Spanish Society of Pathology. *Clinical and Translational Oncology* , 20(7):815–826.

Conlon, A., Taylor, J., and Sargent, D. J. (2014). Multi-state models for colon cancer recurrence and death with a cured fraction. *Statistics in Medicine*, 33(10):1750–1766.

Currie, C. S., Fowler, J. W., Kotiadis, K., Monks, T., Onggo, B. S., Robertson, D. A., and Tako, A. A. (2020). How simulation modelling can help reduce the impact of COVID-19. *Journal of Simulation*, 14(2):83–97.

Dabrowska, D. M. (1987). Nonparametric regression with censored survival time data. *Scandinavian Journal of Statistics*, 14:181–197.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *The Annals of Statistics*, 17:1157–1167.

Daigeler, A., Zmarsly, I., Hirsch, T., Goertz, O., Steinau, H., Lehnhardt, M., and Harati, K. (2014). Long-term outcome after local recurrence of soft tissue sarcoma: a retrospective analysis of factors predictive of survival in 135 patients with locally recurrent soft tissue sarcoma. *British Journal of Cancer*, 110(6):1456–1464.

Duffy, M., Harbeck, N., Nap, M., Molina, R., Nicolini, A., Senkus, E., and Cardoso, F. (2017). Clinical use of biomarkers in breast cancer: Updated guidelines from the

European Group on Tumor Markers (EGTM). *European Journal Cancer*, 75:284–298.

Effraimidis, G. and Dahl, C. M. (2014). Nonparametric estimation of cumulative incidence functions for competing risks data with missing cause of failure. *Statistics & Probability Letters*, 89:1–7.

Efron, B. (1967). The two sample problem with censored data. *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, 4(University of California Press, Berkeley, CA):831–853.

Efron, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association*, 76(374):312–319.

Efron, B. (1994). Missing data, imputation, and the bootstrap. *Journal of the American Statistical Association*, 89(426):463–475.

Földes, A. and Rejto, L. (1981). Strong uniform consistency for nonparametric survival curve estimators from randomly censored data. *The Annals of Statistics*, 9(1):122–129.

Frandsen, B. R. (2019). Testing censoring point independence. *Journal of Business & Economic Statistics*, 37(3):496–505.

Galician Healthcare Service (2021). Dirección Xeral de Saúde Pública. `https://www.sergas.es/Saude-publica`. Online accessed: 06-April-2022.

Gao, J. and Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association*, 103(484):1584–1594.

García-Vicuña, D., Esparza, L., and Mallor, F. (2021). Hospital preparedness during epidemics using simulation: the case of COVID-19. *Central European Journal of Operations Research*, pages 1–37.

Geerdens, C., Janssen, P., and Van Keilegom, I. (2020). Goodness-of-fit test for a parametric survival function with cure fraction. *TEST*, 29(3):768–792.

González-Manteiga, W. and Cadarso-Suárez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *Communications in Statistics Theory and Methods*, 4(1):65–78.

Hanin, L. and Huang, L. (2014). Identifiability of cure models revisited. *Journal of Multivariate Analysis*, 130:261–274.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

Hsu, C., He, Y., Li, Y., Long, Q., and Friese, R. (2016). Doubly robust multiple imputation using kernel-based techniques. *Biometrical Journal*, 58(3):588–606.

Huang, X., Wolfe, R. A., and Hu, C. (2004). A test for informative censoring in clustered survival data. *Statistics in Medicine*, 23(13):2089–2107.

Huang, X. and Zhang, N. (2008). Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach. *Biometrics*, 64(4):1090–1099.

Hurvich, C., Simonoff, J., and Tsai, C. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 60(2):271–293.

Ibrahim, J. G., Chen, M., Sinha, D., Ibrahim, J., and Chen, M. (2001). *Bayesian Survival Analysis*. Springer.

Iglesias-Pérez, M. C. and González-Manteiga, W. (1999). Strong representation of a generalized product-limit estimator for truncated and censored data with some applications. *Journal of Nonparametric Statistics*, 10(3):213–244.

Jackson, D., White, I. R., Seaman, S., Evans, H., Baisley, K., and Carpenter, J. (2014). Relaxing the independent censoring assumption in the Cox proportional hazards model using multiple imputation. *Statistics in Medicine*, 33(27):4681–4694.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data*, volume 1230. Springer.

Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics*, 35(1):139–156.

Laska, E. M. and Meisner, M. J. (1992). Nonparametric estimation and testing in a cure model. *Biometrics*, 48(4):1223–1234.

Lee, S. and Wolfe, R. A. (1998). A simple test for independent censoring under the proportional hazards model. *Biometrics*, 54(3):1176–1182.

Legrand, C. (2021). *Advanced Survival Models.* CRC Press.

Li, C., Taylor, J. M., and Sy, J. P. (2001). Identifiability of cure models. *Statistics & Probability Letters*, 54(4):389–395.

Li, G. and Datta, S. (2001). A bootstrap approach to nonparametric regression for right censored data. *Annals of the Institute of Statistical Mathematics*, 53:708–729.

Li, Q. and Racine, J. S. (2008). Nonparametric estimation of conditional distribution and quantile functions with mixed categorical and continuous data. *Journal of Business Economics & Statistics*, 26(4):423–434.

Liang, H., de Uña-Álvarez, J., and Iglesias-Pérez, M. C. (2012). Asymptotic properties of conditional distribution estimator with truncated, censored and dependent data. *TEST*, 21(4):790–810.

Lin, D., Oakes, D., and Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika*, 85(2):289–298.

Lin, L. and Huang, L. (2019). Connections between cure rates and survival probabilities in proportional hazards models. *Stat*, 8(1):255.

Lindsay, B. G., Markatou, M., and Ray, S. (2014). Kernels, degrees of freedom, and power properties of quadratic distance goodness-of-fit tests. *Journal of the American Statistical Association*, 109(505):395–410.

Lipsitz, S. R., Zhao, L. P., and Molenberghs, G. (1998). A semiparametric method of multiple imputation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):127–144.

López, O. (2009). Single-index regression models with right-censored responses. *Journal of Statistical Planning and Inference*, 139(3):1082–1097.

López, O., Patilea, V., and Van Keilegom, I. (2013). Single index regression models in the presence of censoring depending on the covariates. *Bernoulli*, 19(3):721–747.

López-Cheda, A. (2018). *Nonparametric inference in mixture cure models.* PhD thesis, Universidade da Coruña. `https://ruc.udc.es/dspace/bitstream/handle/2183/20772/LopezCheda_Ana_TD_2018.pdf` [Online: accessed 06-April-2022].

López-Cheda, A., Cao, R., Jácome, M. A., and Van Keilegom, I. (2017a). Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Computational Statistics & Data Analysis*, 105:144–165.

López-Cheda, A., Jácome, M. A., and Cao, R. (2017b). Nonparametric latency estimation for mixture cure models. *TEST*, 26(2):353–376.

López-Cheda, A., Jácome, M. A., Cao, R., and De Salazar, P. M. (2021). Estimating lengths-of-stay of hospitalised COVID-19 patients using a non-parametric model: a case study in Galicia (Spain). *Epidemiology & Infection*, 149(e102).

López-Cheda, A., Jácome, M. A., Van Keilegom, I., and Cao, R. (2020). Nonparametric covariate hypothesis tests for the cure rate in mixture cure models. *Statistics in Medicine*, 39(17):2291–2307.

López-Cheda, A., Jácome, M. A., and López-de-Ullibarri, I. (2021). npcure: An R Package for Nonparametric Inference in Mixture Cure Models. *The R Journal*, 13(1):21–41.

Ma, L., Hu, T., and Sun, J. (2015). Sieve maximum likelihood regression analysis of dependent current status data. *Biometrika*, 102(3):731–738.

Maller, R. A. and Zhou, S. (1992). Estimating the proportion of immunes in a censored sample. *Biometrika*, 79(4):731–739.

Maller, R. A. and Zhou, S. (1994). Testing for sufficient follow-up and outliers in survival data. *Journal of the American Statistical Association*, 89(428):1499–1506.

Martínez-Camblor, P. and de Uña-Álvarez, J. (2013). Studying the bandwidth in k-sample smooth tests. *Computational Statistics*, 28(2):875–892.

Müller, U. U. and Van Keilegom, I. (2019). Goodness-of-fit tests for the cure rate in a mixture cure model. *Biometrika*, 106(1):211–227.

Nadaraya, E. A. (1964a). On estimating regression. *Theory of Probability and Its Applications*, 9(1):141–142.

Nadaraya, E. A. (1964b). Some new estimates for distribution functions. *Theory of Probability & Its Applications*, 9(3):497–500.

Nicolaie, M. A., Taylor, J. M., and Legrand, C. (2019). Vertical modeling: analysis of competing risks data with a cure fraction. *Lifetime Data Analysis*, 25(1):1–25.

Nieto-Baraja, L. E. and Yin, G. (2008). Bayesian semiparametric cure rate model with an unknown threshold. *Scandinavian Journal of Statistics*, 35(3):540–556.

Othus, M., Li, Y., and Tiwari, R. C. (2009). A class of semiparametric mixture cure survival models with dependent censoring. *Journal of the American Statistical Association*, 104(487):1241–1250.

Pan, H., Gray, R., Braybrooke, J., Davies, C., Taylor, C., McGale, P., Peto, R., Pritchard, K. I., Bergh, J., Dowsett, M., et al. (2017). 20-year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. *New England Journal of Medicine*, 377(19):1836–1846.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076.

Patilea, V. and Van Keilegom, I. (2020). A general approach for cure models in survival analysis. *Annals of Statistics*, 48(4):2323–2346.

Peng, Y. and Yu, B. (2021). *Cure models: methods, applications, and implementation.* Chapman and Hall/CRC, Boca Raton, Florida.

Quintela-del Río, A. and Estévez-Pérez, G. (2012). Nonparametric kernel distribution function estimation with kerdiest: an R package for bandwidth choice and applications. *Journal of Statistical Software*, 50:1–21.

Reid, N. (1981). Estimating the median survival time. *Biometrika*, 68(3):601–608.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832–837.

Rubin, D. B. (2004). *Multiple Imputation for Nonresponse in Surveys.* John Wiley & Sons, New York, NY.

Safari, W. C., López-de-Ullibarri, I., and Jácome, M. A. (2021). A product-limit estimator of the conditional survival function when cure status is partially known. *Biometrical Journal*, 63(5):984–1005.

Safari, W. C., López-de-Ullibarri, I., and Jácome, M. A. (2022a). Nonparametric kernel estimation of the probability of cure in a mixture cure model when the cure status is partially observed. First revision in *Statistical Methods in Medical Research*, available in `https://dm.udc.es/preprint/main_paper_cure_rate_Safari_et_al.pdf`. Online accessed: 06-April-2022.

Safari, W. C., López-de-Ullibarri, I., and Jácome, M. A. (2022b). Latency function estimation under the mixture cure model when the cure status is available. Submitted.

Seaman, S. R. and White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, 22(3):278–295.

Seaman, S. R., White, I. R., Copas, A. J., and Li, L. (2012). Combining multiple imputation and inverse-probability weighting. *Biometrics*, 68(1):129–137.

Siannis, F., Copas, J., and Lu, G. (2005). Sensitivity analysis for informative censoring in parametric survival models. *Biostatistics*, 6(1):77–91.

Taylor, J. M. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, 51(3):899–907.

The Cancer Genome Atlas (2021). TCGA Research Network. `https://tcgadata.nci.nih.gov/publications/tcga`. Online accessed: 06-April-2022.

Tristen, H. and Jeffrey, R. (2008). Nonparametric econometrics: the np package. *Journal of Statistical Software*, 27(5):1–32.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72(1):20–22.

Tsodikov, A., Ibrahim, J., and Yakovlev, A. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, 98(464):1063–1078.

Vakulenko-Lagun, B., Mandel, M., and Betensky, R. A. (2020). Inverse probability weighting methods for Cox regression with right-truncated data. *Biometrics*, 76(2):484–495.

Van Keilegom, I. and Veraverbeke, N. (1997). Weak convergence of the bootstrapped conditional Kaplan-Meier process and its quantile process. *Communications in Statistics Theory and Methods*, 26(4):853–869.

Verhasselt, A., Flórez, A., Van Keilegom, I., and Molenberghs, G. (2019). The impact of incomplete data on quantile regression for longitudinal data. *FEB Research Report KBI_1906.*

Wang, L., Du, P., and Liang, H. (2012). Two-component mixture cure rate model with spline estimated nonparametric components. *Biometrics*, 68(3):726–735.

Wang, L., Rotnitzky, A., and Lin, X. (2010). Nonparametric regression with missing outcomes using weighted kernel estimating equations. *Journal of the American Statistical Association*, 105(491):1135–1146.

Wang, Q., Linton, O., and Härdle, W. (2004). Semiparametric regression analysis with missing response at random. *Journal of the American Statistical Association*, 99(466):334–345.

Wang, W. (2003). Nonparametric estimation of the sojourn time distributions for a multipath model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(4):921–935.

Wang, Y., He, S., Zhu, L., and Yuen, K. C. (2007). Asymptotics for a censored generalized linear model with unknown link function. *Probability Theory and Related Fields*, 138(1):235–267.

Watson, G. (1964). Smooth regression analysis. *Indian Journal of Statistics*, 26(4):359–372.

Wei, Y., Ma, Y., and Carroll, R. J. (2012). Multiple imputation in quantile regression. *Biometrika*, 99(2):423–438.

Wu, Y., Lin, Y., Li, C., Lu, S., and Shih, W. J. (2014). Asymptotic efficiency of an exponential cure model when cure information is partially known. *International Journal of Statistics and Probability*, 3(3):1.

Xu, J. and Peng, Y. (2014). Nonparametric cure rate estimation with covariates. *Canadian Journal of Statistics*, 42(1):1–17.

Yakovlev, A. Y., Asselain, B., Bardou, V., Fourquet, A., Hoang, T., Rochefediere, A., and Tsodikov, A. (1993). A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer. *Biométrie et Analyse de Données Spatio-Temporelles*, 12:66–82.